Cosmology and particle theory

Thomas Mohaupt, Universität Jena

July 1, 2005

These are my personal notes for the lectures on cosmology and particle theory in the winter semeser 2004/05 and the summer semester 2005 at the university of Jena. It is not a draft for publication, nor fully worked out lecture notes. This understood, everybody is invited to read and use them.

Contents

1 Introduction **7**

1.1 The cosmological standard model 7

1.1.1 The hot big bang model 7

1.1.2 New developements 8

1.2 A brief history of the universe 9

1.3 Appendix: Units 11

2 Observations **13**

2.1 Sources of radiation 13

2.2 Hubble flux 14

2.3 Methods of observation for distance and mass distribution 15

2.4 Matter distribution 19

2.4.1 Distribution of luminous matter 19

2.4.2 Motion of luminous matter 21

2.4.3 Distribution of elements in the universe 21

2.4.4 Estimate of the age of the universe 23

2.4.5 Indications of dark matter 23

2.5 Cosmic Microwave Background (CMB) 25

2.6 Summary 26

3 Intermezzo: why space-time is dynamic **29**

3.1 Newtonian mechanics 29

3.2 Special relativity 30

3.3 Principle of Equivalence, Locality (Nahewirkung) 31

4 Geometry **33**

4.1 Differential Manifolds 33

4.2 Parallel transport, connection, covariant derivative 35

4.3 Torsion and Curvature 36

4.4 Metric 37

4.5 The relativistic particle 42

4.6 Symmetries of space-time 44

4.7 Maximally symmetric spaces and Space-times 49

4.7.1 Three-dimensional maximally symmetric spaces 49

4.7.2 Higher dimensional spheres and hyperboloids 51
4.7.3 Four-dimensional maximally symmetric space-times . . . 53
4.8 FRW space-times 55
4.9 Hubble law, gravitational redshift, and horizons 58
4.9.1 The Hubble law 58
4.9.2 The gravitational redshift 60
4.9.3 The gravitational redshift (2) 63
4.9.4 Horizons 66

5 Gravity 69
5.1 Actions 69
5.1.1 The action principle for field theories 69
5.1.2 The action principle for pure gravity 70
5.1.3 The action principle for gravity and matter 71
5.1.4 The action for a scalar field 72
5.1.5 The action for the Maxwell field 74
5.1.6 The energy momentum tensor for dust 76
5.1.7 Ideal fluid 78
5.1.8 The Maxwell field as an ideal fluid 79
5.1.9 The cosmological constant as an ideal fluid 79
5.1.10 The scalar field as an ideal fluid 80
5.1.11 Summary: equations of state 81
5.1.12 Energy conditions 81
5.2 Cosmological solutions 82
5.2.1 The Friedman equations 82
5.2.2 Solutions with $k = 0$ and $\Lambda = 0$ 85
5.2.3 Solutions with $\Lambda = 0$ and $k \neq 0$ 89
5.2.4 Solutions with $\Lambda \neq 0$ and $p = 0$ 95
5.2.5 Solutions with $\Lambda \neq 0$, matter and radiation 96
5.2.6 The coincidence problem 97
5.2.7 The relation between distance and redshift 97
5.2.8 The age of the universe 99

6 Particle physics 103
6.1 Standard model 103
6.1.1 Particle content 103
6.1.2 Interactions 104
6.1.3 Neutrino oscillations and neutrino masses 126
6.2 Beyond the standard model 130
6.2.1 Grand unified (gauge) theories 130
6.2.2 Gravity and Superunification 131
6.3 Remarks on the literature 135

7 Thermodynamics **137**
7.1 Overview . . . 137
7.1.1 Conceptual remarks . . . 137
7.1.2 Temperature, energy, expansion . . . 138
7.1.3 Conditions for (approximate) thermodynamic equilibrium 139
7.1.4 Short thermal history of the universe . . . 140
7.2 Thermodynamics . . . 142
7.3 Classical one-particle distributions in curved space-time . . . 145
7.3.1 Distributions . . . 145
7.3.2 Thermodynamics of the radiation dominated universe . . . 147
7.3.3 Thermodynamics of a matter dominated universe . . . 148
7.4 Quantum mechanical one-particle distributions . . . 148
7.5 Thresholds and decoupling . . . 154
7.5.1 Thresholds . . . 154
7.5.2 Equilibrium from gauge interactions . . . 154
7.5.3 Decoupling of a ultra-relativistic species . . . 158
7.5.4 Decoupling of a non-relativistic species . . . 159
7.6 Selected applications . . . 161
7.6.1 The photon background . . . 161
7.6.2 The neutrino background . . . 162
7.6.3 Particle densities . . . 164
7.6.4 Transition from radiation to matter dominance . . . 166
7.6.5 Recombination . . . 166

8 Inflation **171**

A Some formulae **173**
A.1 Units and constants . . . 173
A.2 Cosmological formulae . . . 173
A.3 Some integrals and sums . . . 173
A.3.1 Sums related to the Riemann ζ-function . . . 174
A.3.2 Integrals related to the Γ-function . . . 174
A.3.3 Integrals related to the Gaussian integral . . . 175
A.3.4 Integrals of the form $\int_0^\infty dx \frac{x^m}{e^x - \varepsilon}$ . . . 175

Chapter 1

Introduction

Status of cosmology today. Progress in observational techology, cosmology reaches the precision of lab physics (1990s). Hubble, WMAP. Growing overlap with particle physics (astroparticle physcis, non-accelerator particle physics). Probably the most exciting discipline of modern physics.

1.1 The cosmological standard model

1.1.1 The hot big bang model

Basic properties:

1. The universe is homogenous and isotropic at large scales (> some 100 Mpc). Structures (stars, galaxies, clusters, superclusters, walls, voids and cells) can be treated as perturbations.

2. The universe is in a state of expansion, and started out of a very dense, hot state about (10-20) $\cdot 10^9$y ago.

Supported by the 'three columns':

1. Hubble flux: a universal relation between redshift and disctance. Redshift interpreted as indicator of velocity, implying expansion.

2. CMB = cosmic microwave background (2.7 K). Highly isotropic black body radiation. Interpretation: formed when nuclei and electrons combined into atoms, so that photons decoupled from standard model matter.

3. Distribution of elements.
   75 percent H, 24 percent He. Relative abundance of D and He compared to predictions bases on stellar nucleosynthesis. Explained through primordial nucleosynthesis in the hot early universe.

Theoretical framework:

1. Dynamics of space-time: GR
2. Quantum dynamics of matter: QFT, standard model of particle physics, strong, weak, electromagnetic forces, leptons (neutrinos and electron-like particles) and quarks. Hadrons (Mesons and Baryons) are bound states of quarks.

1.1.2 New developements

1. Universe is flat with a very high precision.
2. Only 5 percent of the energy constent of the universe is standard model matter, 25 percent is non-standard model dark matter (inhomogenously distributed at least at smaller scales), 70 percent is a dark energy (homogenously distributed) which drives the acceleration.
3. Universe has entered 'recently' into a state of accelerated expansion, driven by the dark energy. Indicates existence of a very small cosmological constant, or of a quintessence field.
4. The early universe went through a period of accelerated expansion, 'inflation', during which it grew by a factor $\geq e^{60}$. This explains why it is flat today.

Supportive observations:

1. Measurement of temperature fluctuations in the cosmic microwave background.
2. High redshift supernovae.
3. Dark matter tracing through weak gravitational lensing.

Recent experiments: Hubble, WMAP, ... (Older experiments: COBE, Boomerang)

Cosmological indications of non-standard model physics:

1. Baryogenesis (Matter abundance) cannot be explained within the standard model of particle physics.
2. Existence of exotic dark matter.
3. Existence of dark energy, need for a dynamical mechanism to explain inflation.

Theoretical aspects: Growing need for a quantum theory of gravity, and a unified theory of all particle and interactions. Main fields of activity: loop quantum gravity, string theory. Wide range of speculative ideas: quantum geometry, extra dimensions, branes.

1.2 A brief history of the universe

1. $t \leq 10^{-43}s$. $T = 10^{19}GeV$ in hot big bang model (without inflation).

   Effects of quantum gravity become relevant (presumably).
   $t = 0$ corresponds to a singularity of classical GR. Classical effects (scalar fields violating strong energy condition, extra dimensions) or quantum effects (semiclassical, like corrections to Einsteins equations, or quantum phase of space time geometry, like space-time foam.) or stringy effects (in string theory) might replace this by a bounce, a wormhole, an instanton, a transition from a quantum phase, the collision of branes, etc. No empirical information about this state is available, speculations based on untested theories.

2. $t \simeq 10^{-42}s$.

   Start of inflation. Cold universe starts a period of accelerated expansion during which it grows by a factor $\geq e^{60} \simeq 10^{26}$. This make the univese flat, and most likely it erases all information of the period before. Model dependence: we assume here the minimal version of 'new inflation' with one inflaton field and cold dark matter. Some empirical evidence for inflation from CMB.

3. $t \simeq 10^{-18\pm 6}s$.

   Inflation ends, decelerated expansion. Universe 'reheats' immediately or after some delay to at least $T \sim 1MeV$.

4. $t \simeq 10^{-12}s$, $T \simeq 10^2GeV$.

   Electroweak phase transition. Electroweak interaction goes from symmetric phase (vanishing Higgs mass) to broken phase (massive Higgs particle, massive W and Z particles, short ranged weak interactions. In fact this creates the rest masses of all standard model particles.)
   In 'old inflation' phase transitions like this (or a corresponding GUT phase transitions) we used to create inflation. Did not work in the minimal version, see however 'new old inflation' (locked inflation, trapped inflation).
   The electroweak phase transition might be related to baryogenesis/leptogenesis, i.e., creation of an asymmetry between matter and antimatter. Ingredients: C and CP violation, B number violation, no thermodynamic equilibrium. Today it seems that the standard model cannot account for enough CP violation.
   No direct empirical evidence that this transition happened, that universe ever had this temperature. But one needs to explain baryongenesis/leptogenesis somehow.

5. $t \simeq 10^{-5}s$, $T \simeq 0.3GeV$.

   Quark hadron phase transition/deconfinement-confinement transition.
   Transition from quark gluon plasma to confinement phase where quarks are permanently bound to mesons (Pions, Kaons, ...) and baryons (Nukleons, ...).

   No direct empirical evidence that this transition happened, that universe ever had this temperature.

6. $10^{-2}s \leq t \leq 10^{2}s$, $10MeV \geq T \geq 0.1MeV$.

   Primordial nucleosynthesis of D, He, Li, all in ionized form. Accounds for 75 percent of H vs 24 percent of He, and for Deuterium and Helium abundance.

   Neutrinos decouple from remaining matter (photons, charged leptons, hadrons, nuclei). They should be observable today as an isotropic background with $T_\nu \simeq 1.9K$.

   Era of radiation dominance (most of energy carried by photons and relativistic particles).

7. $t \simeq 10^{11}s$ (10000 y). $T \simeq 10eV$.

   Formation of atoms. Photons decouple from baryonic matter. Observation today as CMB with $T_\gamma \simeq 2.7K$.

   Start of matter dominance (most of the energy sits in massive, non-relativistic particles).

8. $379000y \leq t \leq 200 \cdot 10^{6}y$. (Data taken from WMAP webpage)

   Dark ages between formation of CMB and ignition of first stars.

9. $t \simeq 10^{6}y$. Start of structure formation. Bottom up? Stars $\rightarrow$ larger structures: Stars $t \simeq 200 \cdot 10^{6}y$, oldest galaxies $t \simeq 500 \cdot 10^{6}y$ (??) Largest structures seen today (cells) extend over 100 Mpc. Mechanism of structure formation tied to inflation?
   Reionization. Quasars? Protogalaxies? Oldest known galaxy?

10. $t \simeq 10^{10}y$. $T \simeq$ 2.7K

    Matter dominances ends. Dark energy takes over, and a period of accelerated expansion begins.

1.3 Appendix: Units

Energy–Temperature: 1 GeV = 1.16 $\cdot 10^{13}$K.

1 parsec = 1 pc = 3.26 Ly = $3.086 \cdot 10^{18}$ cm. Corresponds to parallaxe of one angular second.
1 Mpc = 10^6 pc.

Redshift vs distance

$$D = \frac{cz}{H_0} = zh^{-1} 3 \cdot 10^6 Mpc \tag{1.1}$$

or

$$1Mpc \simeq z = \frac{h}{3} 10^{-6} \tag{1.2}$$

Chapter 2

Observations

This is a scetch of the most important observational methods and results.
General reference: [13], i.p. Chapter 1.

2.1 Sources of radiation

1. Electromagnetic radiation

   background vs discrete sources
   continuous (thermic, synchroton) vs discrete (absorption, emission lines)

   (a) Emission: 21 cm H line (hyperfine splitting of atomic ground state)

   (b) Emission: microwaves. Dense cold nebula. Molecular spectra.

   (c) Emission: optic, infrared, cm. Atomic recombination lines.

   (d) Absorption: Lyman-α lines. Neutral H in spectra of remote quasars. Very narrow absorption lines. 'Lyman-α forest.' Seem to be caused by intervening (not ejected) material. Protogalaxies? Lyman series is $n = 1$ series. Ly-α is 121.6 nm, Ly-limit is 91.2 nm.

2. Massive particles: cosmic radiation. Origin unknown. Supernovae?

3. Future: neutrinos? Sol, SN1987a.[1]

4. Future: gravitational waves. (GEO 600, Virgo, Ligo, Lisa).
   GEO 600, see SFB-Transregio.
   Lisa is a satellite based joint project of NASA and ESA, to be launched in 2011.

[1]Neutrino are crucial for (super)novae because they act as buffers for the energy. Without them there would not be an explosion.

2.2 Hubble flux

Interprete observed redshift of remote galaxies as indicator of their motion.

Redshift factor

$$z := \frac{\lambda' - \lambda}{\lambda} \tag{2.1}$$

z = redshift, λ' = observed wavelength, λ = wavelength in rest frame (comparison to lab physics).

Naive (wrong) of observed redshift explanation: Doppler effect.

In SRT, the Doppler effect is explained through the relation of length and time scales between different (global) inertial frames in flat space-time. See [10]. Longitudinal Doppler effect:

$$\lambda' = \sqrt{\frac{c+v}{c-v}}\lambda \tag{2.2}$$

Transversal Doppler effect (time dilation, test of SRT):

$$\lambda' = \frac{\lambda}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{2.3}$$

In general one has an angle dependent effect, which also includes aberration.

Expand the redshift generated by the longitudinal Doppler effect to leading order in $\frac{v}{c}$:

$$z := \frac{\lambda' - \lambda}{\lambda} = \left(1 + \frac{v}{c} + \mathcal{O}(\tfrac{v^2}{c^2})\right) - 1 = \frac{v}{c} + \mathcal{O}(\tfrac{v^2}{c^2}) \tag{2.4}$$

It turns out that this leading order effect coincides with the leading order gravitational redshift, which is the correct explanation. (In our universe, the leading order formula is a good approximation for $z \leq 0.3$).

Hubble law:

$$cz \simeq H_0 D \quad \text{or} \quad v \simeq H_0 D \tag{2.5}$$

where D = proper distance.

$$H_0 = h \cdot 100 \frac{km}{s \cdot Mpc} \tag{2.6}$$

Hubble constant (actually, a function of time: $H_0 = H(t_0)$, where t_0 = 'today'). Insecurity in measurement of H (historical):

$$0.4 \leq h \leq 1 \tag{2.7}$$

Recent [9]

$$h \simeq 0.7 \tag{2.8}$$

Distance in terms of redshift:

$$D = \frac{cz}{H_0} = zh^{-1}3 \cdot 10^6 Mpc \tag{2.9}$$

or

$$1Mpc \simeq z = \frac{h}{3}10^{-6} \tag{2.10}$$

2.3 Methods of observation for distance and mass distribution

Principle: cosmic ladder of scales. Overlapping regimes of validity of methods.

1. Measurement of parallaxis. ≤ 100 pc.
2. Comparison of apparent and absolute luminosity of pulsating stars.
   (a) Cepheiden. Characteristic pulsations in colour, luminosity and radial velocity of stellar atmoshpere. Periods: one hour - 50 days. Resonance effect through a gas close to ionization.
   (b) RR Lyrae stars (periods < 1 day).

   Range of application: inside milky way. Hyaden (40-45 pc), Plejaden (130 pc). Magellan cloud, outside milky way. Local group (30 galaxies, ≤ 1 Mpc, some objects in Virgo cluster (20 Mpc).
3. Emprical rules concerning absolute luminosities
   (a) Empirical rule: the average absolute luminosity of the three brightest red supergiants in a galaxy seems to be universal.
   (b) Distribution of luminosity of spherical star clusters is Gaussian. Maximal luminosity seems to be universal.
   (c) Emprical rule: Tully-Fisher relation. Relation between absolute luminosity of spiral and irregular galaxies and their maximal rotation velocity (luminosity - line width law).
   (d) Empirical rule: Faber-Jackson relation. Relation betwen absolute luminosity and dispersion of star velocities for elliptic galaxies. Extension: combine absolute luminosity and luminiosity per area into 'diameter' D_n. Then D_n-σ relation.

   Empirical rules cannot be derived from fundamental physical laws, but seem to be mutually consistent at (30 per cent level).
   Apply at scales up to 100 Mpc. (Virgo and Coma cluster). At larger distances one often invokes Hubbles law to asign distances [13]. (Model dependent.)

4. Sunyaev-Zeldovich effect. Degradation of the CMB amplitude through Compton scattering of CMB photons off electrons in hot gas clouds. Resulting X-radiation gives gas temperature and density of the cloud. Combined date give linear extension. Comparison to the angular extension gives the distence. (Applies to clusters of galaxies). Example of an 'Angle-diameter distances.'

5. Supernovae. Measure velocity of the explosion material (width of emission lines), *spectrum and apparent luminosity as function of the time, and the maximal luminosity.* Types of SN with characteristic universal behaviour. Type Ia seems to have universal maximal luminosity as a function of the ^{56}Ni which is produced and thrown out. Also characteristic in spectrum: ^{56}Co. Interpretation: Double star, matter transfer from partner onto a white dwarf. Explosion near the Chandrasekhar limit (available roughly fixed energy fixed).
   Standard Candle.

   Farthest Ia SN: $z = 1.7$ (or, naively $5 \cdot 10^6\, h^{-1} Mpc$) [14].

6. Gravitational lensing: bending of light through matter.
   Strong lensing by galaxies and clusters produces multiple images of objects behind them, or Einstein rings.
   Weak lensing (i.p. by dark matter) manifests itself in distortions (Polarization effects). Orientation of elliptic galaxies should be random, but weak lensing introduces a weak alignement [3] p. 122.
   Measuring cluster masses. Dark matter mapping.

   Also: Smearing out of accoustic peaks in the CMB. [12] p. 256.

Example of a very far galaxy: Abell 1835 IR 1916 by ESO. $z \simeq 10$, Distance $13,230 \cdot 10^6$ Ly ($\simeq 4000 Mpc$) Time after BB: $470 \cdot 10^6$ y.

Additional background material: Evolution and final states of stars

Relation between luminosity and colour is encoded in the Hertzsprung-Russel diagram. Stars spend most of their life time one the main line, while burning hydrogen.
Light stars (solar mass and below) stay there for billions of years, while massive states (10 solar masses and above) only live for a few million years.
Time of hydrogen burning: $T_{\text{H-burning}} \simeq \frac{1}{M^2}$. Luminosity: $L \simeq M^3$.

Hydrogen burning: fusion of H into ^4He through p-p cycle and (catalytic) CNO cycle. Produces energy (high energetic photons) which heats the outer regions of the star which produce thermal radiation. Produces neutrinos. Solar neutrinos from the p-p play a role in neutrino physics.

Disgression on neutrino physics:
Discovery of solar neutrinos (HOMESTAKE or Davis experiment) used high-energetic neutrinos from certain branches of the p-p cycle (hep and ^8B neutrinos). Modern experiments (Sage, Gallex) can see neutrinos from all branches. Already the first experiments could distinguis solar neutrinos (which convert Chlor to Argon) from reactor neutrinos which don't. Interpretation: solar neutrinos = neutrinos ≠ antineutrinos = reactor neutrinos. Also since first experiments: missing solar neutrinos, i.e., less neutrinos than expected from theory. Interpretation: neutrino oscillations. Neutrino flavours are associated with the charged leptons: electron, muon, tau. Oscillations occur when mass eigenstates are not eigenstates of the Hamiltonian. Implies non-vanishing neutrino mass. Observation of neutrino oscillations at Super-Kamiokande and SNO measures mass differences of order eV. Masses themselves have not been measured, upper mass bound for electron-neutrino is a few eV.
Once hydrogen burning is over stars become unstable and leave the main line. They become inhomogenous, and the developement depends strongly on mass and matter contents. There is a sequence of contractions which heat up the central temperature until the threshold for a new nuclear fusion process is crossed. After He-burning, there can be C, Ne, O and Si burning, if there is enough mass to contract. The last fusion product is Fe, higher elements have lower binding energies per nucleon so that fusion becomes endothermal.[2] During these periods, stars become giants (or supergiants or subgiants), and go to periods where they pulsate or even suffer a nova or supernova explosion.

The final state depends on the mass remaining M_{final} after all nuclear processes which can balance gravity are exhausted and the star finally collapses.

1. White dwarf.
   For light stars the final state is dense, atomic matter. $\rho \simeq (10^5 \dots 10^6) g/cm^3$. Typical radii are 1 percent of the solar radius.

2. Chandrasekhar limit and neutron stars.
   If $M_{\text{final}} \gtrsim 1.2 \dots 1.4 M_{\text{sol}}$, (1.44??) atomic matter becomes unstable, and the final state is densely packed nuclear matter, a degenerate neutron gas. $\rho \simeq (10^{13} \dots 10^{15}) g/cm^3$. Typical radius is of order 10 km.

3. Oppenheimer-Volkov limit and gravitational collapse.

[2] Heavy elements beyond Fe are formed through neutron capture and subsequent β-decay. The elements produced in stellar nucleosynthesis are released into intragalactic space by (super)novae. This is the raw material for the next generation of stars and for planets, etc.

> If $M_{\text{final}} \gtrsim 2 - 3M_{\text{sol}}$,[3] then nuclear matter is unstable with respect to further compression by gravity. There is no known effect which could yield a stable state: total gravitational collapse.

Supernovae: type Ia, Ib, II supernovae are defined by the (more or less) characteristic time dependence of their luminosities. Type II supernovae are interpreted to result from the collapse of a single star, while type I correspond to double star systems, where the presence of a partner significantly modifies the behavior of the collapsar. Type Ia supernova are believed to occur when a white dwarf accretes mass by cannibalizing its partner. The explosion happens when the Chandrasekhar limit is reached. Thus the mass of the supernovae is roughly fixed, so that they can be used as standard candles. To be precise, certain details like the amount of Ni and Co need to be taken into account. [13, 3].

Total collapse and black holes:
In Newtonian gravity the escape velocity is

$$\frac{v_{\text{escape}}}{c} = \sqrt{\frac{2G_N M}{r}} \tag{2.11}$$

For bodies with $r \leq 2G_N M$ it becomes $\geq c$. This motivated John Michell to speculate about ‘dark stars’ in 1783.

Analysis in GR: if a mass is located inside its associated Schwarzschild radius

$$R_S = 2G_N M \tag{2.12}$$

then the object is surrounded by an event horizon. This is a surface of infinite redshift, $z = \infty$. The object is then causally decoupled from its environment: matter and energy can cross the event horizon only from the outside to the inside, but not back. This motivated Wheeler to call such objects ‘black holes.’

The Schwarzschild radius of earth is $8.8mm$, the one of the sun $2.9km$.

The number of stellar black holes in the galaxy can be estimated from knowledge about stellar evolution. Since one needs a final mass M_{final} of two or three solar masses, the initial mass must be higher. Estimate range from 10 to 30 solar masses, while some claim one needs at least 40.[4] Such heavy stars are relatively rare. If one assume that every star which is heavy enough ends as black hole,

[3] This refers to non-rotating objects, and has a great uncertainty.

[4] There are no stars heavier then about 100 solar masses, because such stars are mechanically unstable and shed mass.

then a black hole forms in milky way roughly every 100 years. Over 10 billion years this makes 10^9 stellar black boles. Their total mass would correspond to less then 1 percent of visible mass of the galaxy.
Since black holes are candidates for dark matter, let us briefly mention other types of black holes. The active galactic centers of remote galaxies are interpreted as hosts of supermassive black holes, with masses $\geq 10^9$ solar masses. Matter falling into such a black hole radiates a major fraction of its mass away (up to some 60 percent for rotating black holes, this is the most efficient known process of energy 'production') before crossing the horizon. It is also believed that our galactic center hosts a 'sleeping' supermassive black hole.[5] Again the mass is of order one percent of the luminous matter of a galaxy.

There might also be primordial black holes, produced from inhomogenities in the early universe. (One does not need stars to produce black holes, they can also form from inhomogenities in dense material, or even from the collision of gravitational waves.) Remark (i): it is not clear whether primordial black holes have formed at all, because the early universe was presumably very homogenous. Remark (ii): Some models in which they form predict masses which are so low that they would have decayed by Hawking radiation by now. But they do not necessarily form in large numbers, and for production in the mostlikely mass ranges, they would have decayed by now through Hawking radiation. Thus they cannot contribute to significantly to dark matter either.

Thus the contribution of black holes to dark matter can be at most a few percent of the luminous mass.
Literature: astronomy [19, 20], black holes [21]

2.4 Matter distribution

2.4.1 Distribution of luminous matter

Luminous matter = Stars, hot gas.

1. Spherical star clusters (Population II, old, Halo) and open star clusters (Population I, new, disc).
2. Galaxies
3. Groups of galaxies.

[5] **Vortrag Kolloquium am 15.11.4!**. It seems that active galactic centers correspond to an early state of galactic evolution which ends when the black hole has swallowed all the material in its reach.

Local group, 30 galaxies, 1 Mpc. Includes: milky way, 2 Magallan clouds, Andromeda M31.
Other local clusters at 1 – 10 Mpc: Sculptor, Ursa Major, Centaurus, Leo
More disctant larger clusters: Virgo, 20 Mpc, 212 galaxies. Coma 90 Mpc

4. Superclusters of galaxies

   Local supercluster or Virgo supercluster. $2 \cdot 10^{15} M_{\text{sol}}$.

   Other superclusters: Coma, Pisces-Cetus, Corona Borealis, Hydra Centaurus.

5. Other very large structures.

   (a) Voids (reduced density, 20 percent of average)

       Great void of extension $60h^{-1}Mpc$ at a distance of $150h^{-1}Mpc$.

   (b) Walls

       Great wall: $60h^{-1}Mpc$ times $170h^{-1}Mpc$. $2 \cdot 10^{16} M_{\text{solar}}$.

   (c) Cells

       Further walls and voids. In total a web-like or cell-like structures at largest scales.
       Note, however: structure depends on the scale of averaging.
       Cell structures of 100-200$h^{-1}Mpc$.

6. Homogenity at last?

   At distances $\geq$ some $100h^{-1}Mpc$ matter seems to be distributed homogenously.

Table 1.1 in [13], p.25 for an overview of structures from star clusters to cell structures.

Object	Extension	Density	Internal Distances	Number of Components	Mass / M_{sol}
Spherical star cluster	10-100 pc	10^2-10^3 $(\mathrm{pc})^{-3}$		10^5-10^7	10^5-10^6
Open star cluster	2-20 pc	10 $(\mathrm{pc})^{-3}$		10-10^3	
Galaxies	1-100 kpc	0.02 h^3 $(\mathrm{Mpc})^3$	3-5 pc	10^{11}	10^6-10^{13}
Clusters	2-10 Mpc	10^{-6} h (Mpc)	1 Mpc	10-10^4	10^{12}-10^{15}
Superclusters	50-150 h^{-1} Mpc	10^{-6} $(\mathrm{Mpc})^{-3}$	25 h^{-1} Mpc	5-50	10^{16}
Voids	25-100 h^{-1} Mpc	0.2 of average	50 h^{-1} Mpc		
Cells	100-200 h^{-1} Mpc		100 h^{-1} Mpc		

2.4.2 Motion of luminous matter

Local particular motions. Actual motions of galaxies can be parametrized as

$$v \simeq H_0 D + v_{\mathrm{part}} \tag{2.13}$$

Examples:
Milky way and andromeda approach (collision in $3 \cdot 10^9$y, c.f. Alastair Reynolds, Redemption Ark).
Milky way moves towards the 'great attractor,' a hypothetical concentration of mass with extra $5 \cdot 10^{16} M_{\mathrm{sol}}$ in a distance of $42 h^{-1} Mpc$. [13], p31.
Actually, the particular motions may be to complicated in order to be explained by one single great attractor. [13], p31.

2.4.3 Distribution of elements in the universe

Mechanisms of nucleosynthesis:

1. Stellar nucleosynthesis

   Stellar nucleosynthesis acts by fusion in the core of stars. Hydrogen burning produces ^{4}He from H by p-p cycle or catalysis (CNO cycle). [13] p.42, [15], figure 9, [16] S. 299. This is followed by He, C, Ne, O, Si burning which created elements up to Ni and Fe. Higher elements can be formed by neutron capture followed by β-decay. [16] Kap. 8.2.

2. Primordial nucleosynthesis

Primordial nucleosynthesis is the production of light elements D,^{3}He, ^{4}He and ^{7}Li in the hot early universe. Higher elements are distroyed by photodisintegration.

3. Spallation in interstellar medium

   Spallation is the production of light elements ^{6}Li, ^{9}Be, ^{10}B, ^{11}B through the splitting of ^{12}C by protons from cosmic radiation.

Methodes of observation: direct measurements in the solar system, solar wind and meteorits. Spectra of other stars (interpretation depends on theoretical knowledge of stellar atmoshperes, star evolution and initial conditions). Outside milky way: measurement of ionized ^{4}He in H-II regions (regions of ionized H close to hot stars, $T \geq 10^4 K$). Since ^{4}He is hard to ionize, there could be more than is concluded from observation.

Some examples:

Mass ratio of deuterium from Lyman-α absorption lines in interstellar medium:

$$\frac{D}{H} = (1.6 \pm 0.15) \cdot 10^{-5} \tag{2.14}$$

Mass ratio for Lithium in stars of population II (spherical star clusters in the halo, 'old'):

$$\frac{^7Li}{H} = (1.2 \pm 0.2) \cdot 10^{-10} \tag{2.15}$$

Mass ratio for Lithium in stars of population I (open star clusters in the disc, 'young(er)'):

$$\frac{^7Li}{H} \simeq 10^{-8} \tag{2.16}$$

(effects of stellar nucleosynthesis).

Overall: 75 percent H, 24 percent Helium, 1 percent others.

Implications of the data:

1. Stellar nucleosynthesis cannot account for the observed deuterium D=^{2}H. Deuterium decays at temperature $T \geq 5 \cdot 10^5 K$.
2. Models of star evolution: only 5 percent of the observed ^{4}He is accounted for by stellar nucleosynthesis.

3. Need spallation to account for observed ^{6}Li, ^{9}Be, ^{10}B, ^{11}B.
4. The observed distribution of elements is in good agreement with primordial nucleosynthesis in the hot big bang model. It is one of the three columns of the model.

2.4.4 Estimate of the age of the universe

1. Radioactive decay

   Earth: $3.7 \cdot 10^9$ y.
   Moon: $4.5 \cdot 10^9$ y.
   Meteorits: $4.57 \cdot 10^9$ y.
   Element distribution in the galaxis: 10-20 $\cdot 10^9$y.

2. Age of spherical star clusters through Hertzsprung-Russel diagram.

   13-18 $\cdot 10^9$ y.

Some years ago, their were indications that the age of the universe, as predicted by the hot big bang model was to small to account for the oldest stars. It seems that this has been resolved now (after dark matter and dark energy have been integrated into the model?!), so that an age of $t \simeq 13.7 \cdot 10^9$y (WMAP) is compatible with all other data.

2.4.5 Indications of dark matter

Dark matter = non-luminous matter.

1. Velocity dispersion of galaxies in clusters of galaxies. Virial theorem, applied to Newtons law:[6]

   $$\overline{E}_{\text{kin}} = \tfrac{1}{2}\overline{E}_{\text{pot}} \tag{2.17}$$

   applied to cluster

   $$\tfrac{1}{2}M\overline{v}^2 = \frac{GM^2}{2r_e} \tag{2.18}$$

   where $\overline{v}$ is the average velocity of galaxies inside the cluster and r_e the effective radius of the cluster. The resulting mass is larger than the one seen in visible matter.

2. Rotation curves of (stars inside) galaxies. Rotation velocity as a function of the disctance from the center $v(r)$. Energy conservation, applied to motion of a distant (test) mass around the mass inside its orbit:

   $$\tfrac{1}{2}mv^2(r) = \frac{GmM(r)}{r^2} \Rightarrow \frac{2GM(r)}{r} = v^2(r) \tag{2.19}$$

[6] Does not apply to larger structures, i.p. to motion of clusters inside superclusters.

Constant mass distribution (as expected for the ends of the spirals if all mass is luminous)

$$v(r) \sim r^{-1/2} \tag{2.20}$$

Observed

$$v(r) \sim \text{const} \tag{2.21}$$

implying

$$M(r) \sim r \tag{2.22}$$

in contradiction to visible mass. Radio measurements (21 cm line) also see larger extension than optical.

Alternative to dark matter: modification of Newton's law / GR. Not appealing theoretically. Concrete proposals seem to be in conflict with other measurements and observations.

Dark matter candidates.

1. Baryonic dark matter (protons, neutrons). MACHOS=massive compact halo objects.
   (a) Brown dwarfs. Jupiter-size almost stars.
   (b) Old white dwarfs.
   (c) Stellar black holes. Masses up to 10^6 solar masses. (Unlikely, according to gravitational lensing of quasars and according to theory: expect 1 black hole forming in milky way in 100 y, total 10^8-10^9, or 1 percent of visible matter.)
2. Non-baryonic standard model matter, in particular massive neutrinos. (Not sufficient).
3. Non-standard model dark matter = exotic particles.
   Can be distinguished from SM dark matter through (i) theoretical reasoning (theory of nucleosynthesis) (ii) observation X-ray emission of galaxies (gives estimate of total baryonic matter).
   Non-SM dark matter candidates: weakly interacting massive particles (WIMPS, WIMPSILLAs).
   Axions. (Strong CP problem of QCD). Supersymmetric partner particles (photinos).

Remark on terminology:

Dark matter = non-luminous matter.
Non-exotic dark matter = standard model matter
Exotic dark matter = non-standard model matter, which only couples gravitationally (or through some new interaction of comparable strength) to standard model matter.

Dark energy = non-luminous, perfectly homogenous and isotropic ‘matter.’ Cosmological constant, homogenous scalar field.

Matter refers to everything ‘on the right hand side of Einstein equations’, i.e., characterizable by an energy-momentum tensor. This includes massless particles and homogenous scalar fields, essientially everything but space-time geometry. In specific context, a different terminology is convenient. For example matter-dominated universe = universe dominated by massive (better: non-relativistic) particles, vs radiation-dominated universe = universe dominated by massless (better: ultra-relativistic) particles. In general, our terminology is such that we subsume ‘radiation’ under ‘matter.’

2.5 Cosmic Microwave Background (CMB)

Isotropic microwave radiation with Planckian distribution (the best known), temperature $T = 2.725K$.
Predicted relict of an early hot period of the universe. Decoupling of photons when atoms forms. Adiabatic expansion cools down the radiation.

Planck distribution:

$$\mu d\nu = \frac{8\pi h}{c^3}\nu^3 d\nu \frac{1}{\exp(\frac{h\nu}{k_B T}) - 1} \tag{2.23}$$

Integration gives Sefan Boltzmann law

$$\mu = a_0 T^4 \tag{2.24}$$

where

$$a_0 = \frac{\pi^2}{15}\frac{k_B^4}{h^3 c^3} \tag{2.25}$$

Motion of earth (sun, milky way, local cluster) relative to the averaged matter distribution should induce an anisotropy.
Parametrization of anisotropy:

$$\frac{\Delta T(\vec{n})}{T_0} = \sum_l \sum_{m=-l}^{l} a_{lm} Y_{lm}(\vec{n}) \tag{2.26}$$

SRT Doppler effect:

$$\nu' = \nu \frac{1 + \frac{v}{c}\cos\Theta}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{2.27}$$

where ν is the frequency in the rest frame. Planck distribution goes to Planck distribution under Lorentz transformation with transformed temperature. Distribution generated in the rest frame of the CMB (defined by the average matter distribution which produced it) has frequency ν and temperature T. The observed distribution has frequency ν', but this can be reinterpreted as a distribution with frequency ν and angle dependent temperature T', by shifting the Doppler effect from the frequency to the temperature: $\frac{\nu'}{T} = \frac{\nu}{T'}$:

$$T' = \frac{\sqrt{1-\frac{v^2}{c^2}}}{1+\frac{v}{c}\cos\Theta} T = (1-\tfrac{v}{c}\cos\Theta)T + \mathcal{O}(\tfrac{v^2}{c^2}) \tag{2.28}$$

The observed dipol anisotropy of the CMB,

$$T' = T(\Theta) = T_0 + T_1 \cos\Theta \tag{2.29}$$

$$\frac{\Delta T}{T} = \frac{T-T_0}{T_0} \simeq 3 \cdot 10^{-3} \tag{2.30}$$

can be interpreted as a particular motion of the earth with $v \simeq 400 km/s$ in the direction of Leo.

After the subtraction of a kinematical quadrupol moment, distortions by discrcete sources etc. there remains an anisotropy of

$$\frac{\Delta T}{T_0} = 6 \cdot 10^{-6} \tag{2.31}$$

Interpretation: primordial density fluctuations of the early hot universe are mapped to temperature fluctuations of the CMB. Seed of structure formation. Problem: inhomogenities observed in the CMB correspond to scales of 3000 Mpc (today), which are much larger than the observed inhomogenities of 100 Mpc.

2.6 Summary

'Concordance model' of current cosmological data.
Fit of WMAP data with ΛCDM model = model with cosmological constant + cold dark matter
'WMAP best fit' refers to Table 3 of [22].

1. Matter contents of the universe. Expressed in terms of dimensionless Ω, where $\Omega = 1$ is a flat universe, the boundary case between a closed ($\Omega > 1$) and an open ($\Omega < 1$) universe.

   Total Mass-energy: $\Omega \simeq 1$. (WMAP best fit $\Omega = 1.02 \pm 0.02$).
   Universe is flat.

   (a) Matter (mass-energy which is distributed inhomogenously at scales below 100 Mpc): $\Omega_{\text{Matter}} \simeq 0.3$.

i. Visible (luminous) matter: $\Omega_L \simeq 0.005$.
ii. Dark standard model matter: $\Omega_{D,SM} \simeq 0.045$.
Total standard model matter: $\Omega_B \simeq 0.05$.
iii. Exotic (non-SM) dar matter: $\Omega_D \simeq 0.25$.

(b) Dark energy = homogenously distributed energy. E.g. cosmological constant, or scalar field (quintessence).
$\Omega_\Lambda \simeq 0.7$.
WMAP best fit: $\Omega_\Lambda = 0.73 \pm 0.04$

2. Hubble constant:

$$H_0 = 0.71 \left\{ \begin{array}{c} +0.04 \\ -0.03 \end{array} \right\} \cdot (9.77813 \cdot 10^9 a)^{-1} \tag{2.32}$$

(WMAP best fit)
or

$$h \simeq 0.7 \tag{2.33}$$

3. Age of the universe: $(13.7 \pm 0.2) \cdot 10^9$y.
WMAP best fit.
Consistent with age of oldes known stars.

4. Value of cosmological constant:

$$\Lambda \simeq 2.6 \cdot 10^{-54} m^{-2} \tag{2.34}$$

or

$$\Lambda L_{\text{Planck}}^2 = 10^{-125} \tag{2.35}$$

where

$$L_{\text{Planck}} = 1.6 \cdot 10^{-35} m \tag{2.36}$$

(why so small? 'Naturalness problem.')

5. Temperature of CMB:

$$T = (2.725 \pm 0.002) K \tag{2.37}$$

Chapter 3

Intermezzo: why space-time is dynamic

3.1 Newtonian mechanics

Mechanics: Principle of covariance. Laws of nature take same form in all inertial systems. Implies: existence of distinguished class of reference systems, where inertial forces are absent. Newtons second law

$$\vec{F} = \frac{d\vec{p}}{dt} = m\frac{d^2}{dt^2}\vec{x} \tag{3.1}$$

Viewed from inertial frame, particles move on straight lines if no forces act.

$$\vec{F} = 0 \Rightarrow \vec{x}(t) = \dot{\vec{x}}(0)t + \vec{x}(0) \tag{3.2}$$

Note: Non-Inertial systems $\leftrightarrow$ inertial forces.

$$\vec{x} = \vec{y} + \vec{s}(t) \Rightarrow \vec{F} = m\ddot{\vec{y}} + m\ddot{\vec{s}} = \vec{F}' + \vec{F}'_{\text{inertial}} \tag{3.3}$$

In particular: $\vec{F} = \vec{0}$, but $\ddot{\vec{s}} \neq 0$ implies $m\ddot{\vec{y}} \neq 0$: in a non-Inertial frame, the motion of a free particle is not a line, but curved.

Covariance group of mechanics = Galilei group

Analogy:
inertial systems $\simeq$ global cartesian coordinate systems
general reference systems $\simeq$ general 'curved' coordinates.
Existence of global inertial systems (global cartesian=flat coordinates systems) $\simeq$ absence of curvature.

3.2 Special relativity

Observation: Maxwell equations are not covariant with respect to Galilei but with respect to Poincare transformation.

Postulats: (1) Covariance, i.e., form invariance of physical equations under change of inertial system. (2) Constancy of speed of light. Implication: Covariance group = Poincare group.

Events

$$x = (x^\mu) = (x^0 = ct, \vec{x} = (x^i)) \tag{3.4}$$

$\mu = 0, 1, 2, 3$, and $i = 1, 2, 3$. We usually set $c = 1$.
Space-time metric

$$\eta = (\eta_{\mu\nu}) = \mathrm{diag}(-1, 1, 1, 1) \tag{3.5}$$

Poincare transformation

$$x \to \Lambda x + a \tag{3.6}$$

Proper time

$$-d\tau^2 = ds^2 = -dt^2 + d\vec{x}^2 \tag{3.7}$$

γ-factor:

$$\frac{d\tau}{dt} = \sqrt{1 - v^2} = \frac{1}{\gamma(v)} \tag{3.8}$$

where $v = |\vec{v}|$ and $\vec{v} = \frac{d}{dt}\vec{x}$.

Four-velocity:

$$u = (u^\mu) = \frac{dx}{d\tau} = (\gamma, \gamma\vec{v}) \tag{3.9}$$

Four-momentum:

$$p = (p^\mu) = mu = (m\gamma, m\gamma\vec{v}) = (E, \vec{p}) \tag{3.10}$$

$$E = m\gamma = \sqrt{m^2 + \vec{p}^2} = m + \frac{\vec{p}^2}{2m} + \cdots \tag{3.11}$$

$$p^2 = -m^2 \quad \text{time-like} \tag{3.12}$$

Relativistic Newton law:

$$f = \dot{p} = m\ddot{x} \tag{3.13}$$

where ‘dot’ referes to τ, and

$$f = (f^\mu) = (f^0 = \vec{F} \cdot \vec{v}\gamma, \vec{F}\gamma) \tag{3.14}$$

is the four-force (space-like).
Free particle:

$$f = 0 \Rightarrow x(\tau) = \dot{x}(0)\tau + x(0) \tag{3.15}$$

time-like straight line in Minkowski space.

Massless particles and light rays:

$$-d\tau^2 = ds^2 = 0 = -dt^2 + d\vec{x}^2 \tag{3.16}$$

Thus $v = |\frac{d\vec{x}}{dt}| = 1$ and there is no proper time (infinite time dilatation).
Momentum of a photon:

$$p = (E, \vec{p}) \tag{3.17}$$

where

$$p^2 = 0 \,, \quad \text{or} \quad E = |\vec{p}| \tag{3.18}$$

Consider a photon moving along the 3-direction:

$$p = (E, 0, 0, E) \tag{3.19}$$

Lightlike curve, viewed from some inertial frame $t, \vec{x}$:

$$x^\mu(t) = (1, 0, 0, 1)\, t \quad \frac{dx^\mu}{dt} = (1, 0, 0, 1) \quad \frac{d^2x^\mu}{dt^2} = 0 \tag{3.20}$$

There is no proper time, but as we see we can use the time t of our inertial frame to parametrize the worldline. The time t is an example of an affine parameter. Observe that the equation

$$\frac{d^2x^\mu}{dt^2} = 0 \tag{3.21}$$

is invariant under affine transformations $t \to at + b$, where $a, b \in \mathbb{R}$. This covers changes between inertial systems, which rescale the time by the (time-independent) γ-factor, and an arbitrary choice of the point $t = 0$. We also see that momentum and velocity are proportional. Thus we can formulate 'Newton's law' for massless particles using an arbitrary affine parameter:

$$\frac{dp}{d\lambda} = f \tag{3.22}$$

In absence of forces this becomes $\frac{dp}{d\lambda} = 0$ which is solved by a lightlike straight line.

3.3 Principle of Equivalence, Locality (Nahewirkung)

Observation: Inertial mass = gravitational mass

Inertial mass m_t:

$$\vec{F} = m_t \ddot{\vec{x}} \tag{3.23}$$

Gravitational mass:

$$\vec{F} = -G_N m_s m_s' \frac{\vec{x} - \vec{x}'}{|\vec{x} - \vec{x}'|^3} = -m_s \vec{\nabla} \phi_{\text{Newton}} \tag{3.24}$$

$$\phi_{\text{Newton}} = -G_N m_s' \frac{1}{|\vec{x} - \vec{x}'|} \tag{3.25}$$

Compare to Coulomb force

$$\vec{F} = \frac{1}{4\pi\varepsilon_0} q q' \frac{\vec{x} - \vec{x}'}{|\vec{x} - \vec{x}'|^3} = -q \vec{\nabla} \phi_{\text{Newton}} \tag{3.26}$$

$$\phi_{\text{Coulomb}} = \frac{q'}{4\pi\varepsilon_0} \frac{1}{|\vec{x} - \vec{x}'|} \tag{3.27}$$

Newton's law is not compatible with special relativity (analogous to going from Coulomb law to Maxwell theory: 'Nahewirkungsprinzip') Elevator thought experiment: gravity is (locally) an inertial force. Inertial systems are locally realized as freely falling systems. Space-time has a (pseudo-) Riemannian metric, which is determined dynamically by the distribution of matter. Conversely, the geometry of space-time determines the (gravitational) motion of matter (stress energy, including not only mass, but all forms of energy).

Minser, Thorne and Wheeler: 'Space-time tells matter how to move, matter tells space-time how to curve.'

Chapter 4

Geometry

4.1 Differential Manifolds

A topological Hausdorff[1] space M is called a real n-dimensional differential manifold, if there exists a C^∞-atlas. This means that

1. for each point $p \in M$ there exists an open neighbourhood U which is homoeomorphic to $\mathbb{R}^n$. The sets U are called the charts of M. The points $x^\mu \in \mathbb{R}^n$ serve as coordinates on M. The collection of all charts is called an atlas of M.

2. on each overlapp of of charts, the composition of maps

$$\mathbb{R}^n \ni x^\mu \to p \to y^\alpha \in \mathbb{R}^n \tag{4.1}$$

   which relates the two systems (x^μ) and (y^α) of coordinates is a C^∞-diffeomorphism. This means that the charts match together into a consistent atlas.

A curve on M:

$$C : t \in I \subset \mathbb{R} \longrightarrow C(t) \in M \tag{4.2}$$

Using coordinates $x = (x^\mu)$:

$$t \to x^\mu(t) \tag{4.3}$$

Tangent vector of C at $t = t_0$:

$$v(t_0) = \left.\frac{dx}{dt}\right|_{t=t_0} \tag{4.4}$$

The tangent vectors of all curves through $p \in M$ span the tangent space

$$T_pM \simeq \mathbb{R}^n \tag{4.5}$$

[1] Hausdorff space means that any two point have dijoint open neighbourhoods.

Vector field:

$$v : p \in M \to v_p \in T_p M \tag{4.6}$$

In components:

$$x \to v(x) = v^\mu(x) e_\mu(x) \tag{4.7}$$

Standard basis of $T_p M$:

$$e_\mu = \left. \frac{\partial}{\partial x^\mu} \right|_p \tag{4.8}$$

Evaluation of vector field on a function = directional derivative:

$$v[f]_p = v^\mu(x) \frac{\partial f}{\partial x^\mu} \tag{4.9}$$

Behaviour under coordinate transformations:

$$x \to y(x) \tag{4.10}$$

induces change of basis in $T_p M$:

$$e_\mu(x) = e_\alpha(y) \frac{\partial y^\alpha}{\partial x^\mu} \tag{4.11}$$

Now

$$v_p = v^\mu(x) e_\mu(x) = v^\alpha(y) e_\alpha(y) \tag{4.12}$$

implies

$$v^\alpha(y) = \frac{\partial y^\alpha}{\partial x^\mu} v^\mu(x) \tag{4.13}$$

This is the old-fashioned definition of a vector field. Kovector field:

$$\omega : p \in M \longrightarrow \omega_p \in T_p^* M \tag{4.14}$$

$$\omega_p = \omega_\mu(x) e^\mu(x) \tag{4.15}$$

where $e^\mu(x) e_\nu(x) = \delta^\mu_\nu$. Standard basis:

$$e^\mu_p = dx^\mu_p \ , \quad dx^\mu(\partial_\nu) = \delta^\mu_\nu \tag{4.16}$$

$$\omega_\alpha(y) = \frac{\partial x^\mu}{\partial y^\alpha} \omega_\mu(x) \tag{4.17}$$

Natural pairing $T_p^* M \times T_p M \to \mathbb{R}$:

$$\langle \omega, \ v \rangle_p = \omega(v)_p = \omega_\mu(x) v^\mu(x) = \omega_\alpha(y) v^\alpha(y) \tag{4.18}$$

Tensor field of type (r, s):

$$t : p \in M \longrightarrow t_p \in (T_p M)^{\otimes r} \otimes (T_p^* M)^{\otimes s} \tag{4.19}$$

$$t^{\alpha_1 \cdots \alpha_r}_{\beta_1 \cdots \beta_s}(y) = \frac{\partial y^{\alpha_1}}{\partial x^{\mu_1}} \cdots \frac{\partial x^{\nu_1}}{\partial y^{\beta_1}} t^{\mu_1 \cdots}_{\nu_1 \cdots}(x) \tag{4.20}$$

4.2 Parallel transport, connection, covariant derivative

Free particle in Minkowski space move have worldlines which are straight lines. This concept admits two generalizations, which coincide for the space-times of Einstein gravity: autoparallel curves and curves of stationary length (geodesics). Here we discuss autoparallel curves.
A priori, we do not know what 'parallel' means. A rule which defines how to 'parallel transport' a vector along curves is called a connection.
Closely related problem: the partial derivative of a tensor field is not a tensor field. As in electrodynamics (or in more general gauge theories) one can define a modified, 'covariant derivative', such that the derivative of a tensor is again a tensor. It turns out (see below) that each such covariant derivative determines a connection (and vice versa). Example: vector field

$$v^\alpha(y) = \frac{\partial y^\alpha}{\partial x^\mu}(x) v^\mu(x) \tag{4.21}$$

$$\partial_\beta v^\alpha(y) = \frac{\partial y^\alpha}{\partial x^\mu}\frac{\partial x^\nu}{\partial y^\beta}\frac{\partial v^\mu}{\partial x^\nu}(x) + \frac{\partial x^\nu}{\partial y^\beta}\frac{\partial^2 y^\alpha}{\partial x^\mu \partial x^\nu} v^\mu(x) \tag{4.22}$$

Covariant derivative

$$\nabla_\nu v^\mu(x) = \partial_\nu v^\mu(x) + \Gamma^\mu_{\nu\rho}(x) v^\rho(x) \tag{4.23}$$

This is a tensor field of type $(1,1)$ iff

$$\Gamma^\alpha_{\beta\gamma}(y) = \frac{\partial y^\alpha}{\partial x^\mu}\frac{\partial x^\nu}{\partial y^\beta}\frac{\partial x^\rho}{\partial y^\gamma}\Gamma^\mu_{\nu\rho}(x) - \frac{\partial x^\nu}{\partial y^\beta}\frac{\partial x^\rho}{\partial y^\gamma}\frac{\partial^2 y^\alpha}{\partial x^\rho \partial x^\nu}(x) \tag{4.24}$$

Analogy: gauge fields in gauge theories are also connections.

The difference of two connections is a tensor field

Covariant derivatives of covector fields:

$$\nabla_\alpha \omega_\beta = \partial_\alpha \omega_\beta - \Gamma^\gamma_{\alpha\beta}\omega_\gamma \tag{4.25}$$

Covariant derivatives of tensor fields:

$$\nabla_\alpha t^{\beta_1\cdots}_{\gamma_1\cdots} = \partial_\alpha t^{\beta_1\cdots}_{\gamma_1\cdots} + \Gamma^{\beta_1}_{\alpha\mu_1} t^{\mu_1\beta_2\cdots}_{\gamma_1\cdots} + \cdots - \Gamma^{\nu_1}\alpha\gamma_1 t^{\beta_1\cdots}_{\nu_1\gamma_2\cdots} - \cdots \tag{4.26}$$

Now we can define what it means that a vector field is parallel along a curve: the directional derivative of the vector field along the curve (i.e., in the direction of the tangent vector) vanishes:

$$\nabla_{\dot{x}} v = 0 \tag{4.27}$$

In components:

$$\dot{x}^{\mu}\nabla_{\mu}v^{\nu} = 0 \tag{4.28}$$

or

$$\dot{v}^{\nu} + \Gamma^{\nu}_{\mu\rho}\dot{x}^{\mu}v^{\rho} = 0 \tag{4.29}$$

One then says that vectors along C are related by parallel transport.

Autoparallel curve: take $v = \dot{x}$, i.e., impose that the tangent vectors are parallel along the curve. 'Curve is generated by parallel transport.' 'Curve is autoparallel.'

$$\dot{x}^{\mu}\nabla_{\mu}\dot{x}^{\nu} = 0 \tag{4.30}$$

or

$$\ddot{x}^{\nu} + \Gamma^{\nu}_{\mu\rho}\dot{x}^{\mu}\dot{x}^{\rho} = 0 \tag{4.31}$$

Not that this is not invariant under general resparametrizations $t \to s(t)$ of the curve, but only under affine reparametrizations $t \to at + b$. The general definition of an autoparallel is therefore

$$\frac{d^2x^{\mu}}{ds^2} + \Gamma^{\mu}_{\nu\rho}\frac{dx^{\nu}}{ds}\frac{dx^{\rho}}{ds} = \lambda(s)\frac{dx^{\mu}}{ds} \tag{4.32}$$

or

$$\nabla_{\dot{x}}\dot{x} = \lambda\dot{x} \tag{4.33}$$

This means that the directional derivative need not to vanish, but only must be parallel to the tangent vector itself. (Loosely speaking the tangent vector may 'changes length'. C.f. normalized vs. non-normalized tangent vectors wrt to non-cartesian coordinates in flat space).
However, one can show that given a geodesic in the general sense, one can always find a reparametrization $s \to t$ such that the geodesic equation takes the form (4.31). Such an 'affine' parametrization is then unique up to affine reparametrizations $t \to at + b$, $a, b \in \mathbb{R}$ constant.

4.3 Torsion and Curvature

Torsion tensor (field) $T^{\rho}_{\mu\nu}$:

$$(\nabla_{\mu}\nabla_{\nu} - \nabla_{\nu}\nabla_{\mu})f = -T^{\rho}_{\mu\nu}\nabla_{\rho}f \tag{4.34}$$

Curvature tensor (field) $R_{\mu\nu\lambda}{}^{\rho}$:

$$(\nabla_{\mu}\nabla_{\nu} - \nabla_{\nu}\nabla_{\mu})v^{\rho} = R_{\mu\nu\lambda}{}^{\rho}v^{\lambda} - T^{\lambda}_{\mu\nu}\nabla_{\lambda}v^{\rho} \tag{4.35}$$

A connection is torsion-free iff $T^{\rho}_{\mu\nu} = 0$. In components:

$$\Gamma^{\rho}_{\mu\nu} = \Gamma^{\rho}_{\nu\mu} \Longleftrightarrow T^{\rho}_{\mu\nu} = 0 \tag{4.36}$$

Torsion measures the non-closure of geodesic parallelogramms.

We will only consider torsion-free connections.

Curvature measures what happens when a vector is parallel-transported around a closed loop (holonomy group $\mathrm{Hol}(\nabla) \subset GL(n, \mathbb{R})$).

Flat connection = Curvature zero.
Flat manifold = exists a flat connection.

4.4 Metric

Local observers can measure length and time intervalls. The corresponding geometrical structure is a metric.

Pseudo-Riemannian metric: field of non-degenerate real symmetric bilinear forms (smooth in $p \in M$).

$$\forall p \in M \; g(\cdot, \cdot)_p : (v_p, w_p) \in T_pM \otimes T_pM \to \mathbb{R} \tag{4.37}$$

in components:

$$g(v, w)_p = g_{\mu\nu}(x) v^\mu(x) w^\nu(x) \tag{4.38}$$

where $(g_{\mu\nu}(x))$ is a non-degenerate real symmetric matrix.

It is useful to think of the tangent space as defining a local inertial frame in which local measurements are performed. The fact that the metric is allowed to vary from point to point reflects that there are no global inertial frames, and that length and time 'units' may vary from point to point.

Space-times of Einstein gravity have a Pseudo-Riemannian metric with signature: $(- + ++)$.

(M, g) is a (Pseudo-)Riemannian manifold.

A curve is called timelike, lightlike (or null), spacelike at a point p, iff

$$g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu < 0 \, , = 0 \, , > 0 \, . \tag{4.39}$$

A curve is called timelike, etc, if it is timelike for all its points.

Volume element:

$$dV = \sqrt{-\det(g_{\mu\nu})}dx^0dx^1dx^2dx^3 =: \sqrt{-g}dx^0dx^1dx^2dx^3 \ . \tag{4.40}$$

Line element:

$$ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu \tag{4.41}$$

Proper length (geodesic length) of spacelike curve:

$$L = \int_{t_1}^{t_2} \sqrt{g_{\mu\nu}(x)\dot{x}^\mu \dot{x}^\nu}dt =: \int \sqrt{g_{\mu\nu}dx^\mu dx^\nu} =: \int dl \tag{4.42}$$

Infinitesimal length element:

$$dl = \sqrt{ds^2} \tag{4.43}$$

The proper time along a timelike curve:

$$T = \int_{t_1}^{t_2} \sqrt{-g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu}dt =: \int \sqrt{-g_{\mu\nu}dx^\mu dx^\nu} =: \int d\tau \tag{4.44}$$

Infinitesimal proper time intervall:

$$d\tau = \sqrt{-ds^2} \tag{4.45}$$

Lightlike curves have length zero. Proper length and proper time are affine parameters. (Lightlike curves have affine parametrizations too.)

A connection ∇ is called compatible with a given metric g, if scalar products of vectors do not change under parallel transport. I.e.

$$\dot{x}^\mu \nabla_\mu g(v, w)_p = 0 \tag{4.46}$$

for all curves x, all parallel vector fields v, w, at all points p. In components

$$\nabla_\mu g_{\nu\rho} = 0 \tag{4.47}$$

If a connection is metric compatible, differentiation is compatible with identifying $T_p^*M \simeq T_pM$ using the metric:

$$v_\mu = g_{\mu\nu}v^\nu \Rightarrow \nabla_\rho v_\mu = g_{\mu\nu}\nabla_\rho v^\nu \tag{4.48}$$

(thus we can talk about covariant and contravariant components of a tensor field.)

Fact: for a given metric, there exists a unique connection, the Levi-Civita connection, which is (i) torison free and (ii) metric compatible.
Components of Levi-Civita connection, Christoffel symbols:

$$\Gamma^\rho_{\mu\nu} = -\tfrac{1}{2}g^{\rho\lambda}(\partial_\lambda g_{\mu\nu} - \partial_\mu g_{\lambda\nu} - \partial_\nu g_{\mu\lambda}) \tag{4.49}$$

Useful formulae[2]

$$\Gamma^{\mu}_{\nu\mu} = \partial_\nu \log\sqrt{-g} \tag{4.50}$$

$$\nabla_\mu v^\mu = \frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}v^\mu) \tag{4.51}$$

$$\nabla_{\mu_1} F^{[\mu_1\cdots\mu_p]} = \frac{1}{\sqrt{-g}}\partial_{\mu_1}(\sqrt{-g}F^{[\mu_1\cdots\mu_p]}) \tag{4.52}$$

Riemann curvature tensor:

$$R_{\mu\nu\rho}{}^{\sigma} = \partial_\nu \Gamma^\sigma_{\mu\rho} - \partial_\mu \Gamma^\sigma_{\nu\rho} + \Gamma^\alpha_{\mu\rho}\Gamma^\sigma_{\alpha\nu} - \Gamma^\beta_{\nu\rho}\Gamma^\sigma_{\beta\mu} \tag{4.53}$$

Symmetry properties:

$$R_{\mu\nu\rho\sigma} = -R_{\nu\mu\rho\sigma} = -R_{\mu\nu\sigma\rho} = R_{\rho\sigma\mu\nu} \tag{4.54}$$

$$R_{[\mu\nu\rho]\sigma} = 0 \tag{4.55}$$

Independent components $\frac{1}{12}n^2(n^2-1)$.

Bianchi identity:

$$\nabla_{[\alpha}R_{\mu\nu]\rho}{}^{\sigma} = 0 \tag{4.56}$$

Remark: this is analogous to the homogenous Maxwell equations $\nabla_{[\mu}F_{\nu\rho]} = 0$ which (locally) imply the existence of a gauge potential A_μ, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. In the case at hand the curvature tensor plays the role of the field strength and the Christoffel symbols the one of the gauge potential.

Flat space-time:

$$R_{\mu\nu\rho\sigma} = 0 \Leftrightarrow g_{\mu\nu} = \eta_{\mu\nu} \tag{4.57}$$

(locally), where $\eta_{\mu\nu} = \text{diag}(-1,1,1,1)$.(Globally one might need more than one coordinate system to cover the manifold. Thus there need not be a globally defined flat coordinate system. Example: tori.)

In general, one can introduce Riemannian normal coordinates wrt a point x_0:

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + \mathcal{O}((x-x_0)^2) \tag{4.58}$$

meaning that

$$\Gamma^\rho_{\mu\nu}(x_0) = 0 \tag{4.59}$$

but in general

$$\partial_\sigma \Gamma^\rho_{\mu\nu}(x_0) \neq 0 \quad \text{and} \quad R_{\mu\nu\rho\sigma}(x_0) \neq 0 \tag{4.60}$$

[2]Straight brackets $[\cdots]$ indicate antisymmetrization. When antisymmetrizing in p indices, we include a combinatorial factor $(p!)^{-1}$.

Remark: if $R_{\mu\nu\rho\sigma}(x_0) = 0$, then this is a coordinates independent statement.

Decomposition of curvature tensor into irreducible tensors (with respect to the reduced structure group $O(p,q) \subset GL(p+q,\mathbb{R})$ of the tangent bundle). Define

1. Ricci tensor:
$$R_{\mu\nu} = R_{\mu\rho\nu}{}^{\rho} \tag{4.61}$$
$$R_{\mu\nu} = R_{\nu\mu} \tag{4.62}$$
2. Ricci scalar:
$$R = R^{\mu}_{\mu} \tag{4.63}$$
3. Weyl tensor (traceless part):
$$C_{\mu\nu\rho\sigma} = R_{\mu\nu\rho\sigma} - \frac{4}{n-2} g_{[\mu[\rho} R_{\sigma]\nu]} + \frac{2}{(D-1)(D-2)} g_{[\mu[\rho} g_{\sigma]\nu]} R \tag{4.64}$$

(Antisymmetrization in p indices includes a combinatorical factor $\frac{1}{p!}$.)

Conformal equivalence:
$$C_{\mu\nu\rho\sigma} = C'_{\mu\nu\rho\sigma} \Leftrightarrow g_{\mu\nu} = e^f g'_{\mu\nu} \tag{4.65}$$
(locally). Conformal flatness:
$$C_{\mu\nu\rho\sigma} = 0 \Leftrightarrow g_{\mu\nu} = e^f \eta_{\mu\nu} \tag{4.66}$$
(locally).
Remark: gravitational waves are excitations of the Weyl tensor.

Flatness in various dimensions:

1. $n = 1$:
$$R_{1111} = 0 \tag{4.67}$$
All one-dimensional manifolds are flat.
2. $n = 2$:
$$R_{1212} \simeq R \tag{4.68}$$
All two-dimensional manifolds are conformally flat.

Einstein tensor ($d = 4$)
$$\mathcal{G}_{\mu\nu} := R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu} \tag{4.69}$$
Bianchi identity implies
$$\nabla^{\mu} \mathcal{G}_{\mu\nu} = 0 \tag{4.70}$$

This is the only covariantly conserved second rank tensor formed out of the curvature which contains no higher derivatives of the metric than the second.[3] It represents the space-time geometry in Einstein's field equation. These equations relate the space-time geometry to the distribution of matter. Matter is represented by the energy momentum tensor $T_{\mu\nu}$, which is conserved in SR: $\partial^\mu T_{\mu\nu} = 0$. In a curved space-time covariance dictates that conservation takes the form $\nabla^\mu T_{\mu\nu}$. The simplest field equation results if on sets a second rank tensor formed out of the curvature proportional to $T_{\mu\nu}$. Consistency then requires that this tensor is also covariantly conserved. If one does not allow higher than second derivatives in the field equations, then the Einstein tensor is the unique choice and the field equation has the form $\mathcal{G}_{\mu\nu} \simeq T_{\mu\nu}$.
A space (space-time) is called an Einstein space, iff

$$\mathcal{G}_{\mu\nu} = \lambda g_{\mu\nu} \Longleftrightarrow R_{\mu\nu} = (\tfrac{1}{2}R + \lambda)g_{\mu\nu} \tag{4.71}$$

Taking the trace on both sides

$$R = (\tfrac{1}{2}R + \lambda)n \tag{4.72}$$

fixes λ in terms of R:

$$\lambda = R(\tfrac{1}{n} - \tfrac{1}{2}) \tag{4.73}$$

Thus

$$R_{\mu\nu} = \tfrac{1}{n} R g_{\mu\nu} \tag{4.74}$$

The Ricci tensor is fixed in terms of the Ricci scalar.

Geodesics

(Metric) geodesic, spacelike: a curve of extremal length.

$$\frac{\delta L}{\delta x^\mu} = 0\,, \quad \text{where} \quad L = \int \sqrt{g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu}\, dt \tag{4.75}$$

Implies

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\rho}\dot{x}^\nu \dot{x}^\rho = \lambda(t)\dot{x}^\mu \tag{4.76}$$

where

$$\lambda(t) = \frac{\ddot{l}(t)}{\dot{l}(t)}\,, \quad l(t) = \int_0^t \sqrt{g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu}\, dt' \tag{4.77}$$

If we take l as the curve parameter:

$$\frac{d^2}{dl^2}x^\mu + \Gamma^\mu_{\nu\rho}\frac{d}{dl}x^\nu \frac{d}{dl}x^\rho = 0 \tag{4.78}$$

since $\frac{dl}{dl} = 1$ and $\frac{d^2 l}{dl^2} = 0$. In this parametrization the tangent vectors are normalized:

$$g_{\mu\nu}\frac{d}{dl}x^\mu \frac{d}{dl}x^\nu = 1 \tag{4.79}$$

[3] The metric itself is also covariantly conserved (even constant). This leads to the possibility of introducing the cosmological constant, to be discussed later.

This is an affine parametrization.

For Riemannian manifolds autoparallel curves (straight curves) and geodescis (curves of extremal lines) coincide.

Space-like curve geodesics have minimal length (locally).

Timelike geodesics:

$$\frac{\delta T}{\delta x^\mu} = 0 \quad T = \int \sqrt{-g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu}dt \tag{4.80}$$

When taking the proper time as parameter we have

$$g_{\mu\nu}\dot{x}^\mu x^\nu = -1 \tag{4.81}$$

and

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\rho}\dot{x}^\nu \dot{x}^\rho = 0 \tag{4.82}$$

Timelike geodesics never have minimal length. Under some conditions they are local maxima (Russel's principle of cosmic lazyness.)

4.5 The relativistic particle

Take test particle to probe the space-time geometry. Test particle means that the feedback of the energy-momentum of the particle on the space-time geometry can be neglected.

Action for a (massive or massless) particle:

$$S = \tfrac{1}{2}\int dt(e(t)^{-1}g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu - e(t)m^2) \tag{4.83}$$

where $x^\mu(t)$ are the coordinates of the particle, t an arbitrary curve parameter, m the mass ($m = 0$ allowed), and $e(t)$ is a non-vanishing ($e(t) \neq 0$ for all t) auxiliary function. This choice of action will be justified a posteriori.

Action principle: the equations of motion are obtained by imposing that the action is stationary with respect to variations of $x^\mu(t)$ and $e(t)$.

1. $$\frac{\delta S}{\delta x^\mu} = 0 \Longrightarrow \ddot{x}^\mu + \Gamma^\mu_{\nu\rho}\dot{x}^\nu \dot{x}^\rho = \frac{\dot{e}}{e}\dot{x}^\mu \tag{4.84}$$

   Geodesic equation in general parametrization. $\dot{e} = 0$ corresponds to an affine parametrization.

2. $$\frac{\delta S}{\delta e} = 0 \Longrightarrow g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu = -e^2 m^2 \tag{4.85}$$

   There are three cases:

(a) $m^2 > 0$. Curve is timelike. Interpretation: worldline of massive particle.

(b) $m^2 = 0$. Curve is null. Interpretation: worldline of massless particle.

(c) $m^2 < 0$. Curve is spacelike. Interpretation: worldline of tachyon, unphysical.

The massive particle. Impose the gauge choice

$$m^2 e^2(t) = 1 \tag{4.86}$$

Then the tangent vector is normalized, $g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu = -1$ and the curve parameter is affine ($\dot{e} = 0$.)
Physics: $\dot{x}^\mu$ is relativistic four-velocity, curve parameter, denoted τ, is the proper time.

Eliminate $e(t)$ from the action through (4.86):

$$e(t) = \frac{1}{m}\sqrt{-g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu} > 0 \tag{4.87}$$

$$S = -m \int d\tau = -m \int \sqrt{-\dot{x}^\mu \dot{x}_\mu} dt \tag{4.88}$$

is the standard action of a relativistic particle.

$$p_\mu = \frac{\partial L}{\partial \dot{x}^\mu} = \frac{m\dot{x}_\mu}{\sqrt{-\dot{x}^\rho \dot{x}_\rho}} \tag{4.89}$$

the four-momentum, satisfies the mass shell condition

$$p^\mu p_\mu = -m^2 \tag{4.90}$$

The equation of motion (with respect to proper time τ)

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\rho}\dot{x}^\nu \dot{x}^\rho = 0 \tag{4.91}$$

reduces for flat space-time $g_{\mu\nu} = \eta_{\mu\nu}$ to the equation of motion of a free relativistic particle

$$\ddot{x}^\mu = 0 \tag{4.92}$$

Therefore, the relativistic Newton law

$$\dot{p}^\mu = m\ddot{x}^\mu = f^\mu \tag{4.93}$$

generalizes for curved space-time as

$$m\left(\ddot{x}^\mu + \Gamma^\mu_{\nu\rho}\dot{x}^\nu \dot{x}^\rho\right) = f^\mu \tag{4.94}$$

where f^μ is the four-vector of non-gravitational forces.

We can also use a non-relativistic parametrization, taking the coordinate time x^0 as curve parameter, $t = x^0$. Then

$$S = -m \int dt \sqrt{1 - \vec{v}^2} \tag{4.95}$$

where $\vec{v} = \frac{d\vec{x}}{dt}$. Spatial part of momentum

$$\vec{p} = \nabla_{\vec{v}} L = \frac{m\vec{v}}{\sqrt{1 - \vec{v}^2}} \tag{4.96}$$

and the energy is given by the Hamiltonian:

$$H = \vec{p}\vec{v} - L = \frac{m}{\sqrt{1 - \vec{v}^2}} = p^0 = E \tag{4.97}$$

The massless particle. For $m^2 = 0$ we cannot eliminate $e(t)$ from the action by a choice of gauge. Nevertheless we can find an affine curve parameter t such that $\dot{e} = 0$. The worldline is a null geodesic, in affine parametrization

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\rho}\dot{x}^\nu\dot{x}^\rho = 0 \tag{4.98}$$

The null-ness of the tangent vector is the mass shell condition of a massless particle

$$p^\mu p_\mu = 0 \tag{4.99}$$

Recall that the four-momentum of a massless particle is

$$p^\mu = (E, \vec{p}) \,, \quad \text{where} \quad E^2 - \vec{p}^2 = 0 \tag{4.100}$$

Spacelike geodescis correspond to particles moving with velocities larger than the speed of light $v > 1$. They are probably unphysical (causality) and formally have negative $m^2 < 0$. Note that the action of a relativistic particle is a useful device for simultanously deriving the geodesic equation for spacelike, timelike and null curves. Note that in the latter case the curve does not have length which could be used to formulate an action principle.

4.6 Symmetries of space-time

Poincare transformations

$$x^\mu \to \Lambda^\mu{}_\nu x^\nu + a^\mu \tag{4.101}$$

are the symmetries of Minkowski space-time

$$ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu \to ds^2 \tag{4.102}$$

We need a generalization of this symmetry concept to curved space-time. More generally we want to define what it means that a tensor field is invariant under a transformation.

Heuristics: a vector field w^μ in $\mathbb{R}^n$ is invariant under translations of the variable (say) x^1, if

$$\partial_1 w^m u(x) = 0 \Leftrightarrow w^\mu = w^\mu(x^0, x^2, x^3, \dots) \tag{4.103}$$

In a curved manifold one can define what it means that a vector field is translation invariant along a curve C. This is expressed in terms of a differential operator, the Lie derivative, which generalizes the notion of a directional derivative. There is one important conceptional point: vectors at different points belong to different tangent spaces. Therefore one does not know how to define meningful differences and differential quotients. We now explain how this is cured.
The first step is to realize that there is a correspondence between a curve and its field of tangent vectors. Given a vector field v^μ we can pose an initial value problem

$$\dot{x}^\mu = v^\mu \quad x^\mu(0) = x_0^\mu \tag{4.104}$$

which locally has a unique solution. Associated with the vector field we have

1. the integral curve
$$C : t \to x^\mu(t) \tag{4.105}$$

2. the one-parameter transformation group
$$g(t) : x_0^\mu \to x^\mu(t) \tag{4.106}$$

3. the flow
$$(t, x_0^\mu) \to x^\mu(t) \tag{4.107}$$

The infinitesimal transformation of a point is given by

$$x_0^\mu \to x_t^\mu = x^\mu(t) = x_0^\mu + t v^\mu(x_0) + \mathcal{O}(t^2) \tag{4.108}$$

Let now $w^\mu(x_0) \in T_{x_0}M$ and $w^\alpha(x_t) \in T_{x_t}M$ be two tangent vectors along the curve. To generalize the directional derivative we need to define their difference. Therefore we need to map $w^\alpha(x_t) \in T_{x_t}M$ back to $T_{x_0}M$. The canonical way to relate these two spaces is the Jacobian of the transformation $x_0 \to x_t$ (and its inverse):

$$\frac{\partial x_t^\alpha}{\partial x_0^\mu} : T_{x_0}M \ni w^\mu(x_0) \to w^\alpha(x_t) = \frac{\partial x_t^\alpha}{\partial x_0^\mu} w^\mu(x_0) \in T_{x_t}M \tag{4.109}$$

$$\frac{\partial x_0^\alpha}{\partial x_t^\mu} : T_{x_t}M \to T_{x_0}M \tag{4.110}$$

Now we can define the Lie derivative of the vector field w^μ along the vector field v^μ at the point x_0:

$$\begin{aligned}
(\mathcal{L}_v w)^\mu(x_0) &= \lim_{t\to 0} \frac{1}{t} \left(\frac{\partial x_0^\mu}{\partial x_t^\alpha} w^\alpha(x_t) - w^\mu(x_0) \right) \\
&= \frac{\partial w^\mu}{\partial x^\nu}(x_0) v^\nu(x_0) - \frac{\partial v^\mu}{\partial x^\nu}(x_0) w^\nu(x_0) \\
&= \left(v^\nu \nabla_\nu w^\mu - w^\nu \nabla_\nu v^\mu\right)(x_0)
\end{aligned} \tag{4.111}$$

A vector field w^μ is invariant along v^μ, if the Lie derivative vanishes, $\mathcal{L}_v w = 0$. In adapted coordinates, where, say, $(v^\mu) = (0, 1, 0, \ldots)$, this reduces to $\partial_1 v^\mu = 0$, which is the definition we used in flat space.
The definition of a Lie derivative can be extended to covariant tensor fields

$$(\mathcal{L}_v T)^{\mu_1 \cdots}(x_0) = \lim_{t \to 0} \tfrac{1}{t} \left(\frac{\partial x^{\mu_1}}{\partial {x_t}^{\alpha_1}} T^{\alpha_1 \cdots}(x_t) - T^{\mu_1 \cdots}(x_0) \right) \quad (4.112)$$

Explicit evaluation:

$$\begin{aligned} (\mathcal{L}_v T)^{\mu_1 \cdots}(x) &= v^\nu \partial_\nu T^{\mu_1 \cdots}(x) - \partial_{\alpha_1} v^{\mu_1} T^{\alpha_1 \mu_2 \cdots}(x) - \cdots && (4.113) \\ &= v^\nu \nabla_\nu T^{\mu_1 \cdots}(x) - \nabla_{\alpha_1} v^{\mu_1} T^{\alpha_1 \mu_2 \cdots}(x) - \cdots && (4.114) \end{aligned}$$

For contravariant tensor fields: use inverse Jacobian, and replace minus-sign by plus-sign.
A tensor field is invariant under the transformation group generated by a vector field v^μ, iff

$$(\mathcal{L}_v T)^{\mu_1 \cdots}(x) = 0 \quad (4.115)$$

Again one can introduce adapted coordinates, etc.
The symmetries of space-time, which generalize the Poincare symmetry of flat Minkowski space, are the isometries of the metric, i.e., transformations which leave the metric invariant:

$$(\mathcal{L}_v g)_{\mu\nu} = \nabla_\mu v_\nu + \nabla_\nu v_\mu = 0 \quad (4.116)$$

(using $\nabla_\mu g_{\nu\rho} = 0$.) A vector field which generates an isometry is called a Killing vector field.
In adapted coordinates, $v^\mu_{(\alpha)} = (\delta^\mu_\alpha)$ the Killing equation becomes

$$\partial_\alpha g_{\mu\nu} = 0 \quad (4.117)$$

i.e. the metric is independent of the coordinate along the integral line of the transformation.

For a Killing vector field one has

$$\nabla_{(\mu} v_{\nu)} = 0 \quad (4.118)$$

and

$$\nabla_\mu \nabla_\nu v_\rho = R_{\rho\nu\mu}{}^\sigma v_\sigma \quad (4.119)$$

Therefore at any point at most

$$v_\mu \quad \text{and} \quad \nabla_{[\mu} v_{\nu]} \quad (4.120)$$

are independent quanitities. The maximal number of Killing vector fields is

$$n + \tfrac{1}{2} n(n-1) = \tfrac{1}{2} n(n+1) \quad (4.121)$$

This bound is saturated for flat space, which is invariant under $O(p,q)\cdot\mathbb{R}^{p+q}$.

For Euclidean space $p = 0, q = 3$ this is the Euclidean group $O(3)\cdot\mathbb{R}^3$, with Killing vector fields

$$\partial_i \quad \text{and} \quad x_i\partial_j - x_j\partial_i \tag{4.122}$$

which generate translations and rotations (for the rotations the integral lines are circles, not straight lines, and the adapted coordinates are angles).

For Minkowski space ($p = 1, q = 3$) this is the Poincare group. The corresponding Killing vector fields are

$$\partial_\mu \quad \text{and} \quad x_\mu\partial_\nu - x_\nu\partial_\mu \tag{4.123}$$

which generate translations and Lorentz transformations (rotations and boosts.)

Spaces (space-times) which have the maximal number of isometries are called *maximally symmetric*. They are not necessarily flat, but have constant curvature:

$$R_{\mu\nu\rho\sigma} = k(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}) \tag{4.124}$$

(k is a constant by Bianchi identity plus metric postulate).

$$R_{\mu\nu} = k(n-1)g_{\mu\nu} \tag{4.125}$$

$$R = k(n-1)n \tag{4.126}$$

and therefore

$$R_{\mu\nu\rho\sigma} = \frac{R}{n(n-1)}(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}) \tag{4.127}$$

The whole curvature tensor is fixed by the Ricci scalar. Note

$$R_{\mu\nu} = \tfrac{1}{n}Rg_{\mu\nu} \tag{4.128}$$

In particular, a maximally symmetric space is an Einstein space.

Maximally symmetric spaces are symmetric (pseudo-)Riemannian spaces, i.e., they are of the form G/K, where G is the group of isometries and K is a maximal subgroup of involutive isometries. Every point of the manifold is the fixed point of an involutive isometry.

A space (space-time) is called *homogenous*, if there exists a group of isometries which acts transitively. I.e. any to points x and $\hat{x}$ can by mapped to one another by an isometry. This means that for every 'direction' α, there must be a Killing vector field which acts as translation. In adapted coordinates, these Killing vectors form a standard basis of T_xM:

$$v^\mu_{(\alpha)} = \delta^\mu_\alpha \tag{4.129}$$

so that

$$x^\mu \to x^\mu + t v^\mu_{(\alpha)} + O(t^2) \tag{4.130}$$

is a translation in the α-direction.
If a space is homogenous, it takes the form $M = G/H$, where G is the group of isometries and $H_x \simeq H$ the isotropy group of a point (H_x is subgroup of G which leaves x invariant. The H_x for all points are isomorphic.)

A space (space-time) is called *isotropic around a point* x_0, if the isotropy group of x_0 acts transitively on the unit (pseudo-)sphere in $T_{x_0}M$. This means that all rotations (and boosts) around x_0 can be generated by Killing vector fields. An infinitesimal (pseudo-)rotation in the $[\alpha\beta]$-plane takes the form

$$x_0^\mu \to x_0^\mu + t\omega^\mu_{[\alpha\beta]\ \nu} x_0^\nu \tag{4.131}$$

where $(\eta\omega)^T = -\eta\omega$ and $\eta = \mathrm{Diag}((-1)^p(+1)^q)$. (In the following we set $\omega_{\mu\nu} = \eta_{\mu\rho}\omega^\rho{}_\nu$.)

The vector field which generates this transformation takes the form

$$v_{[\alpha\beta]} = \delta^\nu_\alpha \delta^\mu_\beta (x_\mu \partial_\nu - x_\nu \partial_\mu) = 2\delta^{[\nu}_\alpha \delta^{\mu]}_\beta x_\nu \partial_\mu \tag{4.132}$$

and therefore has components

$$v_{[\alpha\beta]} = 2\delta^{[\nu}_\alpha \delta^{\mu]}_\beta x_\nu = \omega^{\mu\nu}_{[\alpha\beta]} x_\nu = \omega^\mu_{[\alpha\beta]\nu} x^\nu \tag{4.133}$$

Note that $v_{[\alpha\beta]} = -v_{[\beta\alpha]}$ and $\omega_{[\alpha\beta]} = -\omega_{[\beta\alpha]}$, which reflects that there are only $\frac{1}{2}n(n-1)$ independent (pseudo-)rotations. Above we said that space-time is called isotropic at x_0 if all (pseudo-)rotations are generated by *Killing* vector fields. This means that there must exist $\frac{1}{2}n(n-1)$ independent Killing vector fields $v_{[\alpha\beta]}$, such that (i) $v_{[\alpha\beta]}(x_0) = 0$ ('no translational part') and (ii) $\nabla^\mu v_{[\alpha\beta]\nu}(x_0)$ generate the Lie algebra $so(p,q)$ ('rotational part generates all possible rotations'). Namely, if Killing vector fields with this property exist one can go to adapted coordinates, where they take the form (4.132). In this parametrization it is obvious that $\nabla^\mu v_{[\alpha\beta]\nu}(x_0)$ form the standard basis of $so(p,q)$:

$$\nabla^\mu v_{[\alpha\beta]\nu}(x_0) = \delta^\mu_\alpha \eta_{\nu\beta} - \eta_{\alpha\nu}\delta^\mu_\beta \tag{4.134}$$

Two useful facts:

1. If a space is isotropic at all points, it is homogenous.
2. A space is homogenous and isotropic iff it is maximally symmetric.

4.7 Maximally symmetric spaces and Space-times

Maximally symmetric spaces, i.e., spaces and space-times which have as many isometries as flat space, play a prominent role in cosmology. In this section we discuss the most important examples.

4.7.1 Three-dimensional maximally symmetric spaces

The three-sphere

Consider $S^3 \subset \mathbb{R}^4$ defined by the equation

$$(x^1)^2 + \cdots + (x^4)^2 = r_0^2 \tag{4.135}$$

This is a 'curved' space, embedded into a flat one. Note that $\mathbb{R}^4$ is an auxiliary space, which serves as a tool to construct the differential manifold S^3. (In particular, the fourth dimension does not have any meaning in a cosmological setting.)

The defining equation is invariant under $G = SO(4)$. This is the isometry group of S^3. The isotropy group of a point is $H = SO(3)$, generated by the rotations which have an rotation axis passing through the point. Thus S^3 is maximally symmetric space

$$S^3 \simeq \frac{SO(4)}{SO(3)} \tag{4.136}$$

To obtain an intrinsic description, we solve the constraint (4.135) in terms of independent coordinates on S^n. One possibility are polar coordinates:

$$\begin{aligned} x^1 &= r_0 \cos\chi && (4.137)\\ x^2 &= r_0 \sin\chi \cos\theta && (4.138)\\ x^3 &= r_0 \sin\chi \sin\theta \cos\phi && (4.139)\\ x^4 &= r_0 \sin\chi \sin\theta \sin\phi && (4.140)\\ & && (4.141) \end{aligned}$$

The metric on S^3 is the pull-back of the flat metric on $\mathbb{R}^4$ to (4.135).

$$\begin{aligned} ds^2 &= (dx^1)^2 + (dx^2)^2 + (dx^3)^2 + (dx^4)^2\big|_{S^3} && (4.142)\\ &= r_0^2 \left(d\chi^2 + \sin^2\chi (d\theta^2 + \sin^2\theta d\phi^1)\right) && (4.143)\\ &= r_0^2 (d\chi^2 + \sin^2\chi d\Omega_{(2)}^2) && (4.144) \end{aligned}$$

where $d\Omega_{(2)}^2$ is the line element of the unit two-sphere. The scalar curvature is positive,

$$R \simeq \frac{1}{r_0^2} \tag{4.145}$$

(Exercise: compute the scalar curvature, for example by using Mathematica or Maple.)

Defining a new coordinate

$$r := \sin\chi \tag{4.146}$$

the metric can be rewritten

$$ds^2 = r_0^2\left(\frac{dr^2}{1-r^2} + r^2 d\Omega^2_{(2)}\right) \tag{4.147}$$

(Relation to Hopf fibration $S^1 \to S^3 \to S^2$.) This will be useful later.

Yet another parametrization is obtained by introducing a new radial coordinate $\overline{r}$ by:

$$r = \frac{\overline{r}}{1+\frac{1}{4}\overline{r}^2} \tag{4.148}$$

In the new coordinate one sees that the metric is conformally flat:

$$ds^2 = \frac{r_0^2}{(1+\frac{1}{4}\overline{r}^2)^2}\left(d\overline{r}^2 + \overline{r}^2 d\Omega^2_{(2)}\right) \tag{4.149}$$

The hyperboloid H^3

Consider $H^3 \subset \mathbb{R}^{1,3}$ defined by

$$-(x^1)^2 + (x^2)^2 + (x^4)^2 + (x^4)^2 = -r_0^2 \tag{4.150}$$

This time we have to choose an embedding space with indefinite signature in order to construct the Euclidean manifold we are interested in. Note that again the embedding space has no meaning in its own. In particular, the time-like direction is not the physical time.

The space defined by the equation above is a hyperboloid, invariant under $SO(1,3)$ with isotropy group $SO(3)$, and a maximally symmetric space

$$H^3 = \frac{SO(1,3)}{SO(3)} \tag{4.151}$$

Coordinates on H^3:

$$\begin{aligned} x^1 &= r_0\cosh\chi && (4.152)\\ x^2 &= r_0\sinh\chi\cos\theta && (4.153)\\ x^3 &= r_0\sinh\chi\sin\theta\cos\phi && (4.154)\\ x^4 &= r_0\sinh\chi\sin\theta\sin\phi && (4.155) \end{aligned}$$

Metric:

$$\begin{aligned} ds^2 &= -(dx^1)^2 + (dx^2)^2 + (dx^3)^2 + (dx^4)^2\big|_{H^3} && (4.156)\\ &= r_0^2(d\chi^2 + \sinh^2\chi d\Omega^2_{(2)}) && (4.157)\\ &= r_0^2\left(\frac{dr^2}{1+r^2} + r^2 d\Omega^2_{(2)}\right) && (4.158)\end{aligned}$$

where

$$r := \sinh\chi \tag{4.159}$$

Curvature is negative:

$$R \simeq -\frac{1}{r_0^2} \tag{4.160}$$

All three-dimensional maximally symmetric spaces

By sending r_0 to infinity, or equivalently, by sending the curvature to zero, both S^3 and H^3 approach the flat space $\mathbb{R}^3$. It can be shown that in three Euclidean dimensions every maximally symmetric space is locally isometric to S^3, H^3 or $\mathbb{R}^3$. Thus, it is classified by the sign of the curvature R. Using spherical coordinates for $\mathbb{R}^3$, the line elements can be written in uniform way:

$$ds^2 = r_0^2\left(\frac{dr^2}{1-kr^2} + r^2 d\Omega^2_{(2)}\right) \tag{4.161}$$

where $k = \frac{R}{|R|} = 1, 0, -1$.
These spaces are conformally flat. Set

$$r = \frac{\overline{r}}{1+\frac{1}{4}k\overline{r}^2} \tag{4.162}$$

Then

$$ds^2 = r_0^2\frac{d\overline{r}^2 + \overline{r}^2 d\Omega^2_{(2)}}{(1+\frac{k}{4}\overline{r}^2)^2} \tag{4.163}$$

4.7.2 Higher dimensional spheres and hyperboloids

The n-sphere

Consider $S^n \subset \mathbb{R}^n$ defined by the equation

$$(x^1)^2 + \cdots + (x^{n+1})^2 = r_0^2 \tag{4.164}$$

The defining equation is invariant under $G = SO(n+1)$. This is the isometry group of S^n. The isotropy group of a point is $H = SO(n)$, generated by the rotations which have an rotation axis passing through the point. Thus S^n is maximally symmetric space

$$S^n \simeq \frac{SO(n+1)}{SO(n)} \tag{4.165}$$

To obtain an intrinsic description, we solve the constraint (4.164) in terms of independent coordinates on S^n. One possibility are polar coordinates:

$$\begin{aligned} x^1 &= r_0 \cos\alpha^1 && (4.166) \\ x^2 &= r_0 \sin\alpha^1 \cos\alpha^2 && (4.167) \\ \vdots &= \vdots && (4.168) \\ x^n &= r_0 \sin\alpha^1 \cdots \cos\alpha^n && (4.169) \\ x^{n+1} &= r_0 \sin\alpha^1 \cdots \sin\alpha^n && (4.170) \end{aligned}$$

The metric on S^n is the pull-back of the flat metric on $\mathbb{R}^{n+1}$ to (4.164).

$$\begin{aligned} ds^2 &= \left.\sum_{i=1}^{n+1} (dx^i)^2 \right|_{S^n} && (4.171) \\ &= r_0^2 \left((d\alpha^1)^2 + \sin^2\alpha^1 (d\alpha^2)^2 + \sin^2\alpha^1 \sin^2\alpha^2 (d\alpha^3)^2 + \cdots \right) && (4.172) \end{aligned}$$

The curvature is positive,

$$R \simeq \frac{1}{r_0^2} \qquad (4.173)$$

The hyperboloid H^n

Consider $H^n \subset \mathbb{R}^{1,n}$ defined by

$$-(x^1)^2 + (x^2)^2 + \cdots + (x^{n+1})^2 = -r_0^2 \qquad (4.174)$$

This is a (two-sheeted) hyperboloid, invariant under $SO(1,n)$ with isotropy group $SO(n)$, and a maximally symmetric space

$$H^n = \frac{SO(1,n)}{SO(n)} \qquad (4.175)$$

Curvature is negative:

$$R \simeq -\frac{1}{r_0^2} \qquad (4.176)$$

Flat space

Flat space $\mathbb{R}^n$ can be obtained as the $r \to \infty$ or, equivalently $R \to 0$ limit of S^n or H^n.
This is a Wigner contraction.
As Riemannian symmetric space:

$$\mathbb{R}^n = \frac{ISO(n)}{SO(n)} \qquad (4.177)$$

4.7.3 Four-dimensional maximally symmetric space-times

By a variation of the construction used before, we can construct curved maximally space-times, namely de-Sitter and Anti-de-Sitter space. We will see later that these are vacuum solutions of the Einstein equations with a positive and negative cosmological constant, respectively.

De Sitter space dS^4

Consider the following hyperboloid in $\mathbb{R}^{1,4}$:

$$-(x^1)^2 + (x^2)^2 + \cdots + (x^5)^2 = +r_0^2 \tag{4.178}$$

This is one-sheeted, and as a symmetric space has the form

$$dS^4 = \frac{SO(1,4)}{SO(1,3)} \tag{4.179}$$

Coordinate system:

$$\begin{aligned}
x^1 &= r_0 \sinh t && (4.180)\\
x^2 &= r_0 \cosh t \cos\chi && (4.181)\\
x^3 &= r_0 \cosh t \sin\chi \cos\theta && (4.182)\\
x^4 &= r_0 \cosh t \sin\chi \sin\theta \cos\phi && (4.183)\\
x^5 &= r_0 \cosh t \sin\chi \sin\theta \sin\phi && (4.184)\\
& && (4.185)
\end{aligned}$$

Metric:

$$\begin{aligned}
ds^2 &= -(dx^1)^2 + (dx^2)^2 + (dx^3)^2 + (dx^4)^2 + (dx^5)^2\big|_{dS^4} && (4.186)\\
&= r_0^2 \left[-dt^2 + \cosh^2 t \left(d\chi^2 + \sin^2\chi d\Omega_{(2)}^2\right)\right] && (4.187)\\
&= r_0^2 \left[-dt^2 + \cosh^2 t d\Omega_{(3)}^2\right] && (4.188)
\end{aligned}$$

Contracting and reexpanding three-spheres.
Topology is $\mathbb{R} \times S^3$.
(Four-dimensional) Curvature is positive

$$R \simeq \frac{1}{r_0^2} \tag{4.189}$$

De Sitter space dS^n

Generalization: a hyperboloid in $\mathbb{R}^{1,n}$:

$$-(x^1)^2 + (x^2)^2 + \cdots + (x^{n+1})^2 = +r_0^2 \tag{4.190}$$

This is one-sheeted, and as a symmetric space has the form

$$dS^n = \frac{SO(1,n)}{SO(1,n-1)} \tag{4.191}$$

The metric has Minkowski signature $(1, n-1)$. The curvature is positive,

$$R \simeq \frac{1}{r_0^2} \tag{4.192}$$

Anti de Sitter space AdS^4

$AdS^4 \subset \mathbb{R}^{2,3}$ is defined by

$$-(x^1)^2 - (x^2)^2 + (x^3)^2 + (x^4)^2 + (x^5)^2 = -r_0^2 \tag{4.193}$$

As a symmetric space

$$AdS^4 = \frac{SO(2,3)}{SO(1,3)} \tag{4.194}$$

Coordinates:

$$\begin{aligned} x^1 &= r_0 \cos t \cosh \chi && (4.195)\\ x^2 &= r_0 \sin t \cosh \chi && (4.196)\\ x^3 &= r_0 \sinh \chi \cos \theta && (4.197)\\ x^4 &= r_0 \sinh \chi \sin \theta \cos \phi && (4.198)\\ x^5 &= r_0 \sinh \chi \sin \theta \sin \phi && (4.199) \end{aligned}$$

Metric:

$$\begin{aligned} ds^2 &= -(dx^1)^2 - (dx^2)^2 + (dx^3)^2 + (dx^4)^2 + (dx^5)^2\Big|_{AdS^4} && (4.200)\\ &= r_0^2(-\cosh^2 \chi dt^2 + d\chi^2 + \sinh^2 \chi d\Omega^2_{(2)}) && (4.201) \end{aligned}$$

The sections at fixed time t are hyperboloids H^3. Time is periodic, but one can go to the universal covering space. Role of boundary.

Anti de Sitter space AdS^n

Consider finally $AdS^n \subset \mathbb{R}^{2,n-1}$, defined by:

$$-(x^1)^2 - (x^2)^2 + (x^3)^2 + \cdots + (x^{n+1})^2 = -r_0^2 \tag{4.202}$$

This is a one-sheeted hyperboloid, which carries a metric of Minkowski signature $(1, n-1)$ and has negative curvature

$$R \simeq -\frac{1}{r_0^2} \tag{4.203}$$

As a symmetric space

$$AdS^n = \frac{SO(2,n-1)}{SO(1,n-1)} \tag{4.204}$$

Topology: $S^1 \times \mathbb{R}^{n-1}$. Closed timelike curves. Universal cover: $\mathbb{R}^n$.

Minkowski space

n-dimensional Minkowski space $\mathbb{R}^{1,n-1}$ is the zero curvature limit of dS^n and of AdS^n. Again a Wigner contraction. As a symmetric space:

$$\mathbb{R}^{1,n-1} = \frac{ISO(1,n-1)}{SO(1,n-1)} \tag{4.205}$$

4.8 FRW space-times

Assumptions concerning cosmological space time

1. Plausibility, simplicity and observations suggest the cosmological prinicple: the universe is (at sufficiently large scales) at any time homogenous and isotropic in space.

2. Concerning the matter one adds Weyl's postulat: after averaging over sufficiently long scales matter moves along a congruence of timelike geodesics. There are no intersection and focal points except at a start point (and possibly and end point) in the (finite or infinite) past (and possibly future). (Ideal gas with laminar, not turbulent motion.)

Use this to restrict the form of the metric.

1. The exists a comoving system coordinates, i.e., matter is at rest in these coordinates. The corresponding time coordinates t defines spatial hypersurfaces $t =$ const., which are orthogonal to the worldlines $x^j =$ const.. The metric is therefore hypersurface-orthogonal. In comoving coordiantes $g_{tj} = 0$ and
$$ds^2 = -g_{tt}(t,\vec{x})dt^2 + g_{ij}(t,\vec{x})dx^i dx^j \tag{4.206}$$

2. The world lines are timelike geodesics. This means that non-gravitational interactions are negligible. One can use the proper time along the world lines to introduce a new time coordinate such that
$$ds^2 = -dt^2 + g_{ij}(t,\vec{x})dx^i dx^j \tag{4.207}$$
The time-like coordinate is the cosmological time, the time measured by a comoving, freely falling observer.

3. The Hubble flux is homogenous and isotropic. Therefore the change of spatial sections in time can only be a function of time. I.e. there is a global scale factor which carries the time dependence.
$$ds^2 = -dt^2 + S^2(t)\hat{g}_{ij}(\vec{x})dx^i dx^j \tag{4.208}$$

4. If we assume that the universe is homogenous and isotropic at any given time, the spatial sections are maximally symmetric:
$$ds^2 = -dt^2 + S(t)^2\left(\frac{dr^2}{1-kr^2} + r^2 d\Omega^2_{(2)}\right) \tag{4.209}$$

These are the FRW space-times.

The FRW space-times include the maximally symmetric space-times $\mathbb{R}^{1,3}, dS^4$ as special cases (special choices of $S(t)$.)

A useful reparametrization. Define modified 'radial' coordinate:

$$r = f_k(\chi) = f(\chi) = \begin{cases} \sin\chi & \text{for } k=1 \\ \chi & \text{for } k=0 \\ \sinh\chi & \text{for } k=-1 \end{cases} \tag{4.210}$$

(For $k = 1$ χ is the standard third angular coordinate of S^3. It will be convenient to consider χ as a radial coordinate, but while for $k = 0, -1$ it has infinite range, it actually is an angular (compact) coordinate for $k = 1$.)

Then

$$ds^2 = -dt^2 + S(t)^2(d\chi^2 + f_k(\chi)^2 d\Omega^2_{(2)}) \tag{4.211}$$

The FRW space-times are conformally flat. Define a new time-like coordinate T by

$$dT = \frac{dt}{S(t)} \tag{4.212}$$

Then we see that a FRW metric is conformally equivalent to an 'Einstein universe':[4]

$$ds^2 = S(T)^2 \left(-dT^2 + d\chi^2 + f(\chi) d\Omega^2_{(2)}\right) \tag{4.213}$$

1. For flat spatial sections, $f(\chi) = \chi$, the term in brackets is flat Minkowski space in polar coordinates.

2. For postive spatial curvature, $\sigma(\chi) = \sin\chi$, introduce new coordinates

$$\begin{aligned} \tau &= \tfrac{1}{2}\left(\tan(\tfrac{\tau-\chi}{2}) + \tan(\tfrac{\tau+\chi}{2})\right) & (4.214) \\ \rho &= \tfrac{1}{2}\left(\tan(\tfrac{\tau-\chi}{2}) - \tan(\tfrac{\tau+\chi}{2})\right) & (4.215) \end{aligned}$$

   Then the term in brackets is conformally equivalent to Minkowski space in polar coordinates:

$$-dT^2 + d\chi^2 + \sin^2\chi d\Omega^2_{(2)} = \frac{-d\tau^2 + d\rho^2 + \rho^2 d\Omega^2_{(2)}}{[1 + (\tau-\rho)^2][1 + (\tau+\rho)^2]} \tag{4.216}$$

3. For negative spatial curvaturs, $f(\chi) = \sinh\chi$ just replace tan by tanh.

Remark: In Einstein gravity, conformal flatness means that no propagating gravitational waves are present.

[4]For $k = 1$ the term in brackets on the rhs is known as the Einstein universe.

Perfect cosmological principle and steady state universes

A stronger version of the cosmological principle is obtained when requiring that the universe looks the same to all comoving observers at all times. This is called the perfect cosmological principle and the corresponding cosmologies are called steady state cosmologies. Originally they were proposed as alternatives to Einsteinian gravity and cosmology but later it was found that Einstein gravity admitts a solution (de Sitter solution) which obeys this principle. Steady state cosmologies require exotic physics. To have a steady state despite the expansion of the universe the continuous creation of matter was postulated. When the CMB was discovered steady state cosmologies were abandoned, because they have no explanation for it. (Another point in favour of the hot big bang is nucleosynthesis.)

Local vs global geometry

Imposing local isometries does not determine the geometry globally. The spaces given above are the simplest spaces with the given local isometries. Other spaces with the same local geometry can be obtained by dividing by the action of a discrete group of freely acting isometries. In particular flat and negatively curved spaces can be made compact (tori, Riemann surfaces of genus > 1.)
In general the set of globally defined isometries gets smaller.

Some authors consider such global identifications and look for signatures in the CMB.

De Sitter space as an FRW cosmology with $k = 1$

De Sitter space

$$ds^2 = r_0^2 \left(-dt^2 + \cosh^2 t \, (d\chi^2 + \sin^2 \chi \, d\Omega^2_{(2)}) \right) \tag{4.217}$$

is an FRW cosmology with $k = 1$ and $S(t) = \cosh t$. The spatial sections are three-spheres which contract up to a minimal radius r_0 and then re-expand.

De Sitter space as an FRW cosmolgy with $k = 0$

Introduce the following coordinates on the hyperboloid (4.178):

$$\begin{aligned} t &= r_0 \log \frac{x^1 + x^2}{r_0} \\ x &= \frac{r_0 x^3}{x^1 + x^2} \\ y &= \frac{r_0 x^4}{x^1 + x^2} \\ z &= \frac{r_0 x^5}{x^1 + x^2} \end{aligned} \tag{4.218}$$

Then the metric takes the form first given by de Sitter (1917)

$$ds^2 = -dt^2 + e^{\frac{2t}{r_0}}(dx^2 + dy^2 + dz^2) \tag{4.219}$$

This an FRW cosmology with $k = 0$ and $S(t) = e^{\frac{t}{r_0}}$. The spatial sections are flat three-dimensional spaces which expand exponentially in time t. The metrics (4.217) and (4.219) are locally isometric. However, the coordinates (t, x, y, z) only cover half of the hyperboloid, because they are only defined for $x^1 + x^2 > 0$. Note in particular that although we used the same symbol t, the time variables appearing in (4.217) and (4.219) are different. Therefore both cosmologies are physically distinct, because the spatial section $t = \text{const}$ are tied to a comoving coordinate system in which matter (averaged over sufficiently large scales) is at rest.

The metric (4.219) has two important applications in cosmology, to be discussed later. The first, nowadays considered obsolete, is its interpretation as a steady state universe which obeys the perfect cosmological principle (Bondi and Gold, 1948, Hoyle, 1948). Note that although the metric (4.219) depends explicitly on time t, it is nevertheless invariant under time translations $t \to t + t_0$, if one rescales the spatial variables x, y, z by a factor $e^{-\frac{t_0}{r_0}}$. In other words spatial sections at different t look the same. (In the steady state model this is achieved by assuming the constant creation of matter, so that the energy density of matter stays constant although the universe expands.) The second application of the metric (4.219) is the inflationary scenario, which assumes that the early universe went through a period where the metric had approximately this form. The exponential expansion during this period explains why the universe is flat today, and how the CMB could thermalize. Instead of the constant creation of matter one postulates that the energy density of the universe during this period was dominated by a scalar with suitable properties, called the inflaton field. This will be discussed in detail later.

4.9 Hubble law, gravitational redshift, and horizons

4.9.1 The Hubble law

FRW metric

$$ds^2 = -dt^2 + S(t)^2 \left(\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2_{(2)} \right) = -dt^2 + S(t)^2 (d\chi^2 + f_k(\chi)^2 d\Omega^2_{(2)}) \tag{4.220}$$

where $f_k(\chi) = r(\chi) = \sin\chi, \chi, \sinh\chi$ for $k = 1, 0, -1$.

We consider – without loss of generality – radial light rays. These are characterized by

$$ds^2 = 0 \quad \text{and} \quad d\phi = d\theta = 0 \tag{4.221}$$

This implies

$$\frac{dt}{S(t)} = \pm d\chi = \pm \frac{dr}{\sqrt{1-kr^2}} \tag{4.222}$$

where the sign is '$-$' for ingoing light rays $t_2 > t_1, \chi_2 < \chi_1$. Integration gives

$$\int_{t_1}^{t_2} \frac{dt}{S(t)} = -(\chi_2 - \chi_1) = \chi_1 - \chi_2 \tag{4.223}$$

for ingoing radial light rays.

Let $D(t)$ denote the proper distance (geodesic distance, physical distance) between two points χ_1 and χ_2 with identical values of ϕ, χ at the FRW-time t. Then

$$D(t) = L = \int_C d\lambda \sqrt{g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu} = -S(t) \int_{r_1}^{r_2} \frac{dr}{\sqrt{1-kr^2}} = \tag{4.224}$$

$$-S(t) \int_{\chi_1}^{\chi_2} d\chi = -S(t)(\chi_2 - \chi_1) = S(t)(\chi_1 - \chi_2)$$

(Minus sign is due to Jacobian from curve parameter λ to r, χ. Note that r, χ are decreasing along the curve for ingoing light rays.)

Taking $\chi_1 = \chi$ and $\chi_2 = 0$, i.e. watching an object with comoving coordinate $\chi = \chi_1$ from the origin $\chi_2 = 0$:

$$D(t) = S(t)\chi \tag{4.225}$$

Even if the object is at rest at the comoving coordinate χ, the proper distance depends on time through the scale factor. For objects moving wrt FRW background, χ becomes a function of time,

$$D(t) = S(t)\chi(t) \tag{4.226}$$

The total velocity v_{tot} , the recession velocity v_{rec} and the particular velocity v_{part} are obtained by differentiation wrt FRW time:

$$v_{\text{tot}} = \dot{D} = \dot{S}\chi + S\dot{\chi} = v_{\text{rec}} + v_{\text{part}} \tag{4.227}$$

Particular velocities are the kinds of velocities we are used to from SRT. In particular, the particular velocity of a light beam/photon is always $1 = c$:

$$ds^2 = 0 = -dt^2 + S(t)^2 d\chi^2 \Rightarrow S(t)\dot{\chi} = 1 \tag{4.228}$$

Note, however, that the coordinate velocity χ depends on $S(t)$ and is generically different from c for photons.
Particular velocities are the kind of velocity an observer in a local inertial frame measures. They are bounded by c.

The recession velocity is of different nature. It characterizes the expansion of space, not the motion of objects in space. It is observable in the sense that it can be inferred from observation (e.g. from the gravitational redshift), but it does not correspond to a velocity of an object measured by a local observer (these are particular velocities). While particular velocities are bounded by c, recession velocities can be arbitrarily large. But this just means that the distance between two points grows at a certain rate, it does not mean that a local observer measures objects passing him with velocities larger than c. In particular, local measurements of the speed of light always yield c (as this is particular motion.) The conceptual mistake in interpretating recession velocities larger than c as being in conflict with special relativity is to assume (implicitly) that there is a global system of observers which are mutually at rest. But this need not be possible in a general FRW space-time.

The total velocity of a photon travelling towards us from a remote galaxy is

$$v_{\mathrm{tot}}(t) = v_{\mathrm{rec}}(t) - 1 \tag{4.229}$$

This can be positive or negative. If the recession velocity is large enough, i.e., if space expands fast enough, then a light ray send to us from a distant galaxy will never reach us. This is related to the notions of event and particle horizons, to be discussed below. Note that the recession velocity can change in time.

Hubbles law, GR version:

$$v_{\mathrm{rec}}(t) = \dot{S}(t)\chi = H(t)D(t) \tag{4.230}$$

where the expansion rate

$$H(t) := \frac{\dot{S}}{S} \tag{4.231}$$

is the Hubble function.

4.9.2 The gravitational redshift

In this section we do not set $c = 1$.

Consider a source at $\chi = 0$ emitting light of wave length λ_1. Let t_1 and $t_1 + \delta t_1$ correspond to two successive peaks, i.e., $\lambda_1 = c\delta t_1$.

Let the signals be absorbed by a detector at rest at $\chi = \chi_2$ at times t_2 and $t_2 + \delta t_2$. Thus the wavelength is $\lambda_2 = c\delta t_2$.

Then

$$\chi = \chi_2 - \chi_1 = c\int_{t_1}^{t_1+\delta t_1} \frac{dt}{S(t)} = c\int_{t_2}^{t_2+\delta t_2} \frac{dt}{S(t)} \tag{4.232}$$

This implies (expand, higher order terms do not contribute, exact result):

$$\frac{\delta t_2}{S(t_2)} = \frac{\delta t_1}{S(t_1)} \tag{4.233}$$

or

$$\frac{\lambda_1}{\lambda_2} = \frac{S(t_1)}{S(t_2)} = \frac{\omega_2}{\omega_1} \tag{4.234}$$

This is the gravitational redshift (better: frequency shift). It reflects that the speed of clocks depends on the space-time metric and changes from point to point (time dilatation). The effect is similar to, but different from the redshift caused by the SRT doppler effect. Here one compares clocks which are at rest in different global inertial frames. In GR there are i.g. no global inertial frames. But (see next section) both the gravitational redshift and the SR doppler effect are limiting cases of a general GR formula.

Define redshift factor

$$z = \frac{\lambda_2 - \lambda_1}{\lambda_1} \tag{4.235}$$

Then

$$1 + z = \frac{S(t_2)}{S(t_1)} \tag{4.236}$$

Note that expansion $S(t_2) > S(t_1)$ corresponds to a redshift $z > 0$, or $\lambda_2 > \lambda_1$, while contraction corresponds to a blueshift. (Since we live in an expanding universe we will use the term redshift instead of the more appropriate term 'frequency shift.')

Relation of redshift to proper distance and recession velocity. Consider an ingoing radial photon emitted at $(t = t_1, \chi = \chi_1)$ and registered at $(t = t_2, \chi = \chi_2 = 0)$. Then

$$\chi = c \int_{t_1}^{t_2} \frac{dt}{S(t)} \tag{4.237}$$

Goal: relate the emission time $t = t_1$ to the redshift z observed at (t_2, χ_2). Define an auxiliary function $z(t)$:

$$1 + z(t) = \frac{S(t_2)}{S(t)} \tag{4.238}$$

so that

$$z(t_1) = z \quad z(t_2) = 0 \tag{4.239}$$

Interpretation: variation of the time of emission within the interval $t_1 \leq t \leq t_2$. Differentiate:

$$\dot{z} = -S(t_2)\frac{\dot{S}(t)}{S(t)^2} = -S(t_2)\frac{H(t)}{S(t)} \tag{4.240}$$

Express the Hubble function in terms of z, $t = t(z)$ and regroup

$$\frac{dt}{S(t)} = -\frac{1}{S(t_2)}\frac{dz}{H(z)} \tag{4.241}$$

Integrate:

$$\int_{t_1}^{t_2} \frac{dt}{S(t)} = -\frac{1}{S(t_2)}\int_z^0 \frac{dz}{H(z)} \tag{4.242}$$

We obtain a elation between the coordinate-distance χ and the redshift z:

$$\chi = c\int_{t_1}^{t_2} \frac{dt}{S(t)} = \frac{c}{S(t_2)}\int_0^z \frac{dz}{H(z)} \neq \frac{c}{S(t_2)}(t_2 - t_1) \tag{4.243}$$

The geodesic distance is

$$D(t_2) = S(t_2)\chi = c\int_0^z \frac{dz}{H(z)} \; . \tag{4.244}$$

Note that the observed redshift is a direct measure for the geodesic distance. By direct we mean that there is no explicit time dependence, because we do not need to know $S(t_2)$, but only $H(z)$. But since we need to integrate over z, we need to know the expansion rate at all times. The consequences are more obvious if we write

$$D(t_2) = cS(t_2)\int_{t_1}^{t_2} \frac{dt}{S(t)} \neq c(t_2 - t_1) \tag{4.245}$$

Here we indicated that the naive estimate $c(t_2 - t_1)$ (valid in a static space-time) for the distance is not accurate if $S(t)$ varies strongly with time.
We can also express the recession velocity as a function of the redshift:

$$v_{\rm rec}(t_2) = c\dot{S}(t_2)\chi = cH(t_2)\int_0^z \frac{dz}{H(z)} \tag{4.246}$$

Note that we need to know $H(t_2)$, i.e., there is an explicit dependence on time. At fixed time the redshift determines the recession velocity, but the relation involves a time-dependent factor of proportionality. Again it is useful to rewrite this as

$$v_{\rm rec}(t_2) = c\dot{S}(t_2)\chi = c\dot{S}(t_2)\int_{t_1}^{t_2} \frac{dt}{S(t)} \neq H(t_2)c(t_2 - t_1) \tag{4.247}$$

Note that the naive estimate $Hc\delta t$ for the recession velocity (valid in a universe with constant expansion rate) can be highly misleading if the Hubble function varies strongly with time.
If z is not to large (rule of thumb: $z \leq 0.3$), $H(z)$ is approximately constant and the integral can be computed approximately:

$$v_{\rm rec} = cz \tag{4.248}$$

Incidentially, the SR Doppler effect has the same leading order approximation. The longitudinal Doppler effect

$$\lambda_2 = \lambda_1 \sqrt{\frac{c+v}{c-v}} \tag{4.249}$$

can be expressed in terms of z as

$$v = c\frac{(1+z)^2 - 1}{(1+z)^2 + 1} \tag{4.250}$$

which implies $v \simeq cz$ for small z (equivalently, small $\frac{v}{c}$).

For large z, the GR FRW redshift and the SR Doppler redshift give different results. It should be clear that both effects have quite different origins. The redshift formula applies to recession velocities in an FRW universe, while the Doppler formula applies to particular velocities. In fact, both effect can be derived from one single more general GR formula, by specializing it to recession and particular velocities, respectively (see next section).
Observation of course always shows a superposition of both effects, and disentangling them is only possible if we decompose the full metric into an FRW background and local perturbation corresponding to (say) galaxies. For example, the blueshift in the spectrum of Andromeda indicates its disctance to us decreases. The interpretation within FRW cosmology is that its particular motion towards us overcompensates FRW recession.
Also note that the FRW recession does not manifest itself in growing distances between stars in a galaxy, etc. Galaxies are gravitational bound states (at a length scale of some 100 kpc) while the FRW recession manifests itself at much larger distances (above some 100 Mpc), where we can approximate matter by a gas whose 'atoms' are the largest gravitationally bound structures (superclusters), and then approximate this gas by a continuum model, the ideal fluid, to be discussed soon. At this level the FRW recession means that the density of the fluid decreases. Going back to the 'atomic' picture this means that the average distance between the largest gravitationally bound structures grows.

4.9.3 The gravitational redshift (2)

Reference: [17, 18].

Consider the motion of a photon in an arbitrary space-time. The photon moves along null geodesic. If we use an affine parameter l, the geodesic equation takes the form

$$k^\mu \nabla_\mu k^\nu = 0 \tag{4.251}$$

where $k^\mu = \frac{dx^\mu}{dl}$. In this parametrization, the tangent vector agrees with the momentum vector, up to constant (and factors of c, which we have set to unity). The affine form of the geodesic equation is the appropriate generalization of

Newton's law, $\dot{k}^\mu = 0$. Since the particle is massless, the affine parameter is not the proper time. The affine parameter is unique up to affine transformations, and the additive constant is irrelevant for the tangent vector. The multiplicative constant is fixed as follows: let u^μ be the four-velocity along the timelike geodesic of an observer who intercepts the photon and measures its energy. By generalization of the SR rule, the energy of the photon is $\omega = k^\mu u_\mu$. This fixes the remaining ambiguity in the relation between momentum and tangent vector in an affine parametrization.
To measure the change of energy of a photon moving along a null geodesic, we take two observers along timelike geodesics with four-velocities u_i^μ, $i = 1, 2$ who measure the momenta k_i^μ of the photon. k_1 and k_2 are related by the geodesic equation (since the vectors k_i are related by parallel transport along the trajectory of the photon). The ratio between the energies is

$$\frac{\omega_1}{\omega_2} = \frac{(k^\mu u_\mu)_1}{(k^\mu u_\mu)_2} \tag{4.252}$$

This formula includes both the gravitational redshift and the Doppler effect. The later is encoded in the velocities of the observers, who need not be at fixed comoving coordinates.
Consider now a FRW space-time in coordinates (t, χ, θ, ϕ). We observe that the only non-vanishing first derivative of the metric is

$$\partial_t g_{\chi\chi} = 2S(t)\dot{S}(t) \tag{4.253}$$

Therefore the only non-vanishing components of the connection are

$$\Gamma^t_{\chi\chi} = S(t)\dot{S}(t)\ , \quad \Gamma^\chi_{t\chi} = \Gamma^\chi_{\chi t} = \frac{\dot{S}(t)}{S(t)} \tag{4.254}$$

Solving the geodesic equation for a radially moving photon gives

$$k^\mu = \left(\frac{1}{S(t)}, \frac{1}{S^2(t)}, 0, 0\right) \tag{4.255}$$

Consider now two observers at rest at points χ_i which measure the photon at times t_i. Their four-velocities have the form $u_i = (1, 0, 0, 0)$. Thus

$$\frac{\omega_1}{\omega_2} = \frac{S(t_2)}{S(t_1)} \tag{4.256}$$

which is the gravitational redshift for FRW geometries. Note that this can is a specialization of the general formula where the particular velocity of the observer vanishes, so that the whole effect is due to the recession velocity corresponding to the time dependence of the FRW metric.

Remark: the FRW metric is invariant under shifts $\chi \to \chi + a$, where a is constant. In the language of Hamiltonian mechanics, the coordinate is cyclic. Therefore there is a corresponding conserved quantity, the canonical momentum p_χ. One can show [17] p. 233, [18] p. 98 that this conserved canonical momentum is

$$p_\chi = S(t)k^0 = S(t)\frac{d\chi}{dt} \tag{4.257}$$

From the redshift formula we can read off the equivalent statement that $\omega S(t)$ is conserved.

Observe that the conserved, canonical momentum and the kinetic momentum differ by the scale factor $S(t)$. In flat space this would be momentum or (equivalent for a massless particle) energy conservation. In FRW cosmologies we have a modified conservation law.

Exercise: show that for flat space and arbitrary u_i one obtains the SRT Doppler formula.
Answer: take

$$k = (1,1,0,0)\ , \quad u_1 = (1,0,0,0,)\ , \quad u_2 = \gamma(v)(1,v,0,0) \tag{4.258}$$

then

$$\frac{ku_2}{ku_1} = \sqrt{\frac{1-v}{1+v}} \tag{4.259}$$

This is the longitudinal Doppler effect. Take

$$u_2' = \gamma(v)(1,0,1,0) \tag{4.260}$$

then

$$\frac{ku_2'}{ku_1} = \gamma(v) = \frac{1}{\sqrt{1-v^2}} \tag{4.261}$$

This is the transverse Doppler effect.
More generally one can obtain a general relative motion an derive the general Doppler effect and aberration. (When comparing to [10], p. 70, take into account aberration. Since transversality is not preserved under Lorentz boosts, there are two possible notions of transverse Doppler effect. Our transverse Doppler effect corresponds to $\cos\overline{\Theta} = 0$ in [10], who discuss the other type of transverse Doppler effect. The relative factor γ between frequencies gets inverted when going from one transversality condition to the other.)

Additional literature

Stephani: go from wave to ray optics. [17] p. 81.

4.9.4 Horizons

Event horizons.

Consider an observer at $\chi = 0$ and sources at $\chi = \chi_1 > 0$ which send radial incoming light rays. The event horizon of the observer at time t is defined by a signal which reaches him only at time t_∞, the latest time, which can be finite or $+\infty$.

$$\chi_{\text{EH}}(t) = \int_t^{t_\infty} \frac{dt}{S(t)} \tag{4.262}$$

Note that a priori $\chi_{\text{EH}}(t) \leq \infty$. If it is finite, there are events 'beyond the horizon' which we will never know about. All event horizons are bounded by the backward light cone at time t_∞.

Example 1: Minkowski space $S(t) = 1$, $k = 0$.

$$\chi_{\text{EH}}(t) = \int_t^{\infty} dt' = \infty \tag{4.263}$$

Minkowski space has no eveent horizons.

Example 2: de Sitter universe (steady state version of de Sitter). $S(t) = e^{\frac{t}{a}}$ where $a > 0$. This has flat spatial sections $k = 0$, thus $\chi = r$ is the standard radial coordinate of flat 3-space.

$$\chi_{\text{EH}}(t) = \int_t^{\infty} dt' e^{-\frac{t'}{a}} = ae^{-\frac{t}{a}} \tag{4.264}$$

This goes to zero at late times, reflecting that the universe expands so fast that the visible part decreases with time.

Particle horizons

Consider the 'opposite' situation. Again we have an observer at $\chi = 0$ and sources at $\chi > 0$. Let t_0 denote the first moment of time, which can be finite or $-\infty$. Then the particle horizon of the observer at the time t is the place from which a signal sent at the time t_0 just reaches him.

$$\chi_{\text{PH}}(t) = \int_{t_0}^{t} \frac{dt'}{S(t')} \tag{4.265}$$

Again, a priori $\chi_{\text{PH}}(t) \leq \infty$. All particle horizons are bounded of by the forward light cone at t_0.

Example 1: Minkowski space

$$\chi_{\text{PH}}(t) = \int_{-\infty}^{t} dt' = \infty \tag{4.266}$$

Minkowski space has no particle horizons.
Example 2a: de Sitter universe

$$\chi_{\mathrm{PH}}(t) = \int_{-\infty}^{t} dt' e^{-\frac{t'}{a}} = \infty \tag{4.267}$$

There is no particle horizon (but an event horizon as we saw above).

Example 2b: full de Sitter space, $S(t) = \cosh \frac{t}{r_0}$. (Note that Minkowski space arises in the limit $r_0 \to \infty$, since $S(t) \to 1$.)

$$\chi_{\mathrm{PH}}(t) = \int_{-\infty}^{t} \frac{dt'}{\cosh \frac{t'}{r_0}} = 2r_0 \arctan e^{\frac{t}{r_0}} \tag{4.268}$$

There is a particle horizon, which is bounded from above by πr_0 even in the limit $t \to \infty$. In the Minkowski limit the particle horizon becomes infinite for all t.

Chapter 5

Gravity

5.1 Actions

5.1.1 The action principle for field theories

Field theory: generalized coordinates are themselves functions of space-time. Denote fields collectively by $\Phi = \Phi(x)$. Action:

$$S[\Phi] = \int d^4x \mathcal{L}(\Phi, \partial\Phi) \tag{5.1}$$

where the Lagrangian $\mathcal{L}$ is a function of Φ and its derivatives. The equations of motion are determined by imposing that $S[\Phi]$ be stationary with respect to variations

$$\Phi \to \Phi + \delta\Phi \ . \tag{5.2}$$

Expand:

$$S[\Phi + \delta\Phi] = S[\Phi] + \int d^4x \left(\frac{\partial\mathcal{L}}{\partial\Phi}\delta\Phi + \frac{\partial\mathcal{L}}{\partial\partial_\mu\Phi}\delta\partial_\mu\Phi \right) + \mathcal{O}[(\partial\Phi)^2] \, . \tag{5.3}$$

Use $\delta\partial_\mu\Phi = \partial_\mu\delta\Phi$, perform partial integration, take boundary terms to vanish:

$$S[\Phi + \delta\Phi] = S[\Phi] + \int d^4x \left(\frac{\partial\mathcal{L}}{\partial\Phi} - \partial_\mu \frac{\partial\mathcal{L}}{\partial\partial_\mu\Phi} \right) \delta\Phi + \mathcal{O}[(\partial\Phi)^2] \, . \tag{5.4}$$

S is stationary, $\frac{\delta S}{\delta\Phi} = 0$, iff the Euler-Lagrange equations

$$\frac{\partial\mathcal{L}}{\partial\Phi} - \partial_\mu \frac{\partial\mathcal{L}}{\partial\partial_\mu\Phi} = 0 \tag{5.5}$$

hold. (In practise, when determining the equations of motion for a given action, it might be more convenient to vary $S[\Phi]$ then to substitute into (5.5).

5.1.2 The action principle for pure gravity

Systematic search for an action principle for gravity, i.e., for the metric. Impose reparametrization invariance. Expansion in the number of derivatives:

$$S_G = \int d^4x \sqrt{-g} \left(\lambda + \tfrac{1}{2\kappa^2} R + \alpha_1 R^2 + \text{other 4-deriv. terms} \right) \tag{5.6}$$

Dimensions:

$$[g_{\mu\nu}] = 0 \,, \quad [d^4x] = -4 \,, \quad [R] = 2 \,, \quad [\kappa] = -1 \tag{5.7}$$

The minimal choice is to restrict to terms with up to two derivatives. All known fundamental actions in physics are of this form. Remark: higher order terms are suppressed at low energies. It might be possible that what we know believe to be fundamental theories are just low energy effective field theories of some other underlying theory. We will restrict ourselves to actions with up to two derivatives.

Variational formulae:

$$\delta g^{\mu\nu} = -g^{\mu\alpha} g^{\nu\beta} \delta g_{\alpha\beta} \tag{5.8}$$

$$\delta \det(g_{\mu\nu}) = \det(g_{\mu\nu}) g^{\alpha\beta} \delta g_{\alpha\beta} \tag{5.9}$$

$$\delta \sqrt{-g} = -\tfrac{1}{2} \sqrt{-g} g_{\alpha\beta} \delta g^{\alpha\beta} \tag{5.10}$$

$$\delta \Gamma^{\mu}_{\nu\rho} = -\tfrac{1}{2} g^{\mu\sigma} (\partial_\sigma \delta g_{\nu\rho} - \partial_\nu \delta g_{\sigma\rho} - \partial_\rho \delta g_{\nu\sigma}) \tag{5.11}$$

$$\sqrt{-g} g^{\mu\nu} \delta R_{\mu\nu} = \partial_\rho (g^{\mu\nu} \delta \Gamma^{\rho}_{\mu\nu} - g^{\mu\rho} \delta \Gamma^{\nu}_{\mu\nu}) \tag{5.12}$$

Variation of the gravitational action:

$$\delta S_G = \int d^4x \sqrt{-g} (\tfrac{1}{2\kappa^2} (R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu}) \tag{5.13}$$

where we defined

$$\Lambda = -\kappa^2 \lambda \tag{5.14}$$

(cosmological constant). Euler-Lagrange equations are the vacuum Einstein equations with cosmological term:

$$R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = 0 \tag{5.15}$$

('Vaccum' means that no matter is present, where 'matter' means any degrees of freedom except those of space-time/gravity itself.)

Solutions of (5.15) are called Einstein spaces. For maximally symmetric solutions,

$$R_{\mu\nu\rho\sigma} = \tfrac{1}{12} R \left(g_{\mu\rho} g_{\nu\sigma} - g_{\mu\sigma} g_{\nu\rho} \right) \tag{5.16}$$

we have

$$C_{\mu\nu\rho\sigma} = 0 \,, \quad R_{\mu\nu} - \tfrac{1}{4} R g_{\mu\nu} = 0 \,, \quad R = \text{const.} \,. \tag{5.17}$$

The maximally symmetric solutions of (5.15) are Minkowski space (for $\Lambda = 0$), de Sitter space (for $\Lambda > 0$) and anti-de Sitter space (for $\Lambda < 0$).

5.1.3 The action principle for gravity and matter

In classical physics, matter is described by either fields or particles.[1] In this section we will consider fields. Let us denote the matter fields collectively by Φ. The action for matter coupled to gravity is is obtained from the matter action in Minkowski space by replacing the Minkowksi metric by the dynamical, Riemannian metric, and by replacing partial derivatives by covariant derivatives. This is dictated by the equivalence principle and by the requirement of reparametrization invariance. Thus the matter action depends on both Φ and $g_{\mu\nu}$. Schematically:

$$S_M[g_{\mu\nu}, \Phi] = \int d^4x \mathcal{L}_M(g_{\mu\nu}, \Phi) \tag{5.18}$$

Examples will be given soon. While the variation with respect to Φ gives the matter equations of motion in curved space-time, the gravitational field equations (which determine the space-time metric) are obtained by the variation of the combined action of gravity and matter:

$$\frac{\delta S_G[g_{\mu\nu}]}{\delta g_{\mu\nu}} + \frac{\delta S_M[g_{\mu\nu}\Phi]}{\delta g_{\mu\nu}} = 0 \tag{5.19}$$

Consider the effect of varying the metric on S_M:

$$\begin{aligned} \delta S_M &= \int d^4x \sqrt{-g} \frac{1}{\sqrt{-g}} \frac{\partial \mathcal{L}_M}{\partial g^{\mu\mu}} \delta g^{\mu\nu} \\ &= \int d^4x \sqrt{-g} \left(-\frac{1}{2} T_{\mu\nu} \right) \delta g^{\mu\nu} \end{aligned} \tag{5.20}$$

where

$$T_{\mu\nu} = \frac{-2}{\sqrt{-g}} \frac{\partial \mathcal{L}_M}{\partial g^{\mu\mu}} \delta g^{\mu\nu} \tag{5.21}$$

is called the energy momentum tensor of the action S_M. Combining this with δS_G we get the full Einstein equations (with cosmological term):

$$R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa^2 T_{\mu\nu} \tag{5.22}$$

It can be shown (for example, by putting a non-relativistic probe particle into an asymptotically flat space-time) that

$$\kappa^2 = 8\pi G_N \tag{5.23}$$

where G_N is Newton's constant. A famous textbook summarizes the content of (5.22) as follows [6]: 'Space acts on matter, telling it how to move. In turn, matter reacts back on space, telling it how to curve.'

[1] In quantum theory and quantum fields theory, particles and fields turn out to correspond to be just different types of states. But we will restrict ourselves to classical physics.

Observe that the space-time curvature enters the Einstein equations in the particular combination

$$\mathcal{G}_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu} \tag{5.24}$$

which is called the Einstein tensor. This is the only object, except the metric itself, which contains up to second derivatives of the metric and is covariantly conserved,

$$\nabla^\mu G_{\mu\nu} = 0 \; . \tag{5.25}$$

Thus the covariant version $\nabla^\mu T_{\mu\nu} = 0$ of energy conservation dictates the structure of the graviational side of the Einstein equations. Note that this local version of energy momentum conservation does not imply the existence of global conserved quantities, which can be interpreted as the mass (total energy in rest frame) and momentum of a space-time. This is, however possible, if the space-time is asymptotically flat and thus represents an 'isolated system' (ADM construction).
Let us some now consider some important matter actions explicitly.

5.1.4 The action for a scalar field

Action:

$$S = \int d^4x \sqrt{-g} \left(-\tfrac{1}{2} g^{\mu\nu} \nabla_\mu \phi \nabla_\nu \phi - V(\phi) \right) \tag{5.26}$$

where the potential $V(\phi)$ is some function of ϕ, which does not depend on the derivatives of ϕ. Not that for a scalar field

$$\nabla_\mu \phi = \partial_\mu \phi \; , \quad \text{but} \quad \nabla_\mu \partial_\mu \phi \neq \partial_\mu \partial_\nu \phi \; . \tag{5.27}$$

(Careful) Variation of scalar field yields its equation of motion:

$$\Box \phi = - \frac{\partial V}{\partial \phi} \tag{5.28}$$

where

$$\Box \phi = g^{\mu\nu} \nabla_\mu \nabla_\nu \tag{5.29}$$

is the (curved space) wave operator or d'Alembertian. A free (non-interacting) scalar field with mass m corresponds to the quadratic potential

$$V = \tfrac{1}{2} m^2 \phi^2 \; . \tag{5.30}$$

The corresponding Euler-Lagrange equation is the (curved space) Klein-Gordon equation

$$(\Box - m^2)\phi = 0 \tag{5.31}$$

Variation of the metric yields energy momentum tensor

$$T_{\mu\nu} = \nabla_\mu \phi \nabla_\nu \phi + g_{\mu\nu} (-\tfrac{1}{2} g^{\alpha\beta} \nabla_\alpha \phi \nabla_\beta \phi - V(\phi)) \tag{5.32}$$

which goes into the Einstein equation. By the equivalence principle the interpretation of its components carries over from flat space physics (special relativity). Thus

$$T_{00} = \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}(\vec{\nabla}\phi)^2 + V(\phi) = \rho(\phi) \tag{5.33}$$

is the energy density (measured by a freely falling observer intercepting the event $x = (t, \vec{x})$,

$$T_{0i} = \dot{\phi}\partial_i\phi \tag{5.34}$$

is the energy flux and

$$T_{ii} = \partial_i\phi\partial_i\phi + \frac{1}{2}\dot{\phi}^2 - \frac{1}{2}(\vec{\nabla}\phi)^2 - V(\phi) \tag{5.35}$$

$$T_{ij} = \partial_i\phi\partial_j\phi \ , \quad i \neq j \tag{5.36}$$

are the components of the stress tensor.

For a spatially homogenous field, $\partial_i\phi = 0$, we have

$$T_{00} = \frac{1}{2}\dot{\phi}^2 + V(\phi) = \rho(\phi) \tag{5.37}$$

and

$$T_{ii} = \frac{1}{2}\dot{\phi}^2 - V(\phi) = p(\phi) \tag{5.38}$$

while $T_{0i} = 0 = T_{ij}$.
As we will see later, this is the energy momentum tensor of a homogenous ideal fluid with energy dentsity ρ and pressure p.

The scalar potential

Let us discuss the physical meaning of the scalar potential. Expand around $\phi = 0$:

$$V = V_0 + g_1\phi + \frac{1}{2}g_2\phi^2 + \frac{1}{3!}\phi^3 + \cdots \tag{5.39}$$

The ground state or vacuum of the scalar field theory is given by the minimum of the scalar potential $\phi = \phi_* = \text{const}$, where $V(\phi_*) = \text{Min}$.

1. V_0 does not enter the equations of motion. But changing V_0 changes the energy of the ground state. If $\phi_* = 0$, then V_0 is the energy of the ground state. Note that the energy of the ground state (also called vacuum energy) is relevant in theories with gravity, because energy is the source of space-time curvature. We will see that a non-vanishing vacuum energy acts like a non-vanishing cosmological constant.

2. g_1 is relevant for the value of ϕ_*, or vacuum expectation value. If $g_1 \neq 0$, then the ground state cannot be $\phi = 0$, because $V'(\phi = 0) = g_1$ while at the minimum we must have $V'(\phi_\star) = 0$, $V''(\phi_*) > 0$.

   If $\phi_* \neq 0$, the field 'has a non-vanishing vacuum expectation value.' One can define a new field $\varphi = \phi - \phi_*$, which has a vanishing vacuum expectation value. Note that when expanding the potential in terms of φ instead of ϕ, the coefficients V_0, g_i change.

3. If the minimum of the potential is at $\phi_* = 0$, then g_2 is the mass-squared of the field (and of the associated particle), $g_2 = m^2$, as we saw above. But in general the mass is determined by the quadratic fluctuations around the minimum of the potential:

$$m^2 = \frac{\partial^2 V}{\partial \phi^2}_{|\phi=\phi_*} = \frac{\partial^2 V}{\partial \phi^2}_{|\varphi=0} \tag{5.40}$$

   Thus we have $g_2 = m^2$ only if the we use the field φ where the vacuum expectation value has been absorbed.

   For a general stationary point $V'(\phi_*) = 0$, we might have $V''(\phi_*) > 0$ (Minimum), $V''(\phi_*) = 0$ (Saddle point) or $V''(\phi_*) < 0$ (Maximum). In the first two cases the theory describes a massive and a massless particle, respectively. A maximum does not correspond to a particle. A field mode sitting at this point is unstable and will move towards a nearby minimum (if any) under any small perturbation. To find the particle interpretation of a scalar field theory one has to find the minima (and saddle points) of the potential. This is important for understanding the Higgs mechanism.

4. Terms of higher than second order in the potential lead to non-linear field equations. These terms therefore desribe interactions. Denoting the interaction part of the potential by

$$V_I = \frac{1}{3!}\phi^3 + \cdots \tag{5.41}$$

   the field equation resembles a generalized Newton law, with the gradient of V_I as force term:

$$(\Box - m^2)\phi = -\frac{\partial V_I}{\partial \phi} \tag{5.42}$$

5.1.5 The action for the Maxwell field

Action:

$$S[g_{\mu\nu}, A_\mu] = \int d^4x \sqrt{-g} \left(-\tfrac{1}{4} F_{\mu\alpha} F_{\nu\beta} g^{\mu\nu} g^{\alpha\beta}\right) \tag{5.43}$$

Field strength:

$$F_{\mu\nu} := \partial_\mu A_\nu - \partial_\nu A_\mu = \nabla_\mu A_\nu - \nabla_\nu A_\mu \tag{5.44}$$

In the second step we used that the Christoffel connection is metric compatible, hence symmetric in its lower indices. The dynamical variables which have to varied are A_μ, not $F_{\mu\nu}$. Variation of the gauge potential gives one half of the vacuum Maxwell equations.

$$\nabla_\nu F^{\mu\nu} = \frac{1}{\sqrt{-g}} \partial_\nu (\sqrt{-g} F^{\mu\nu}) = 0 \tag{5.45}$$

Adding charged matter by minimal coupling. the action becomes:

$$S_M[g_{\mu\nu}, A_\mu] = \int d^4x\sqrt{-g}\left(-\tfrac{1}{4}F_{\mu\alpha}F_{\nu\beta}g^{\mu\nu}g^{\alpha\beta} + A_\mu j^\mu\right) \tag{5.46}$$

and we get the (curved space) inhomogenous Maxwell equations

$$\nabla_\nu F^{\mu\nu} = \frac{1}{\sqrt{-g}}\partial_\nu(\sqrt{-g}F^{\mu\nu}) = j^\mu \tag{5.47}$$

where j^μ is the current describing the charge distribution. (To describe the dynamics of the charge matter we would have to add its action. Possible choices are a complex scalar field or a Dirac field.)
The homogenous Maxwell equations are not Euler Lagrange equations but Bianchi identities (integrability conditions for (5.44))

$$\partial_{[\mu}F_{\nu\rho]} = 0 \tag{5.48}$$

which ensure (locally) the existence of the potential A_μ.
Equivalently (using again the symmetry of the Christoffel connection)

$$\nabla_{[\mu}F_{\nu\rho]} = 0 \tag{5.49}$$

Variation of metric:

$$\delta S = \int d^4x\sqrt{-g}\delta g^{\mu\nu}\left(\tfrac{1}{8}F_{\alpha\beta}F^{\alpha\beta}g_{\mu\nu} - \tfrac{1}{2}F_{\mu\alpha}F_{\nu\beta}g^{\alpha\beta}\right) \tag{5.50}$$

Energy momentum tensor:

$$T_{\mu\nu} = F_{\mu\alpha}F_{\nu\beta}g^{\alpha\beta} - \tfrac{1}{4}F_{\alpha\beta}F^{\alpha\beta}g_{\mu\nu} \tag{5.51}$$

Traceless (off shell):

$$g^{\mu\nu}T_{\mu\nu} = 0 \tag{5.52}$$

Interpretation of the components:

$$F_{0i} = E_i\ , \quad F_{jk} = \epsilon_{jki}B_i \tag{5.53}$$

$$F_{\alpha\beta}F^{\alpha\beta} = 2(\vec{B}^2 - \vec{E}^2) \tag{5.54}$$

$$T_{00} = \tfrac{1}{2}(\vec{E}^2 + \vec{B}^2) = \rho \tag{5.55}$$

is the energy density,

$$T_{0i} = E_j\epsilon_{ikl}B_l\delta^{jk} = (\vec{E}\times\vec{B})_i = \vec{S}_i \tag{5.56}$$

is the Poynting vector (energy flux),

$$T_{ij} = -E_iE_j + \tfrac{1}{2}\vec{E}^2\delta_{ij} + B_iB_j - \tfrac{1}{2}\vec{B}^2\delta_{ij} = -T_{ij}^{\text{Maxwell}} \tag{5.57}$$

is the Maxwell stress tensor (up to sign).

At a given point in space-time, we can diagonalize:

$$T^{\mu\nu} = \mathrm{diag}(\rho, p_1, p_2, p_3) \tag{5.58}$$

For a homogenous field this holds globally. For an isotropic field we have $p_1 = p_2 = p_3 = p$. Tracelessness implies

$$p = \tfrac{1}{3}\rho \tag{5.59}$$

As we will see later this is the equation of state for gas of massless particles and $T_{\mu\nu} = \mathrm{diag}(\rho, p, p, p)$ is the energy momentum tensor for isotropic radiation (massless particles).

5.1.6 The energy momentum tensor for dust

We now turn to the description of matter in terms of particles. When we considered particles before, we used a single particle to probe the space-time geometry and neglected the backreaction. This is called a test particle. Now we want to consider particles as sources of gravitation. Naturally, we then need to consider many particle systems, and then it is convenient to use a 'coarse grained' continuum description. We will not give an action principle, but specify the energy momentum tensor and the equation of motion, now called equation of state. The continuum models will be refered to as dust, gas or fluid. We use fluid as the generic term in the following. The constituents of the fluid can be (clusters of) galaxies, dustparticles, molecules, atoms or elementary particles, depending on context. In the contemporary universe the constituents are the largest existing gravitational bound states, i.e., superclusters. In the hot early universe the constituents are elementary particles.
A relativistic fluid is characterized by a field of four-velocities, $u^\mu(x)$. These are the normalized tangent vectors of the world lines of the infinitesimal elements of the fluid. We will be assumed that the fluid is laminar, so that we have a congruence of world lines (world lines do not cross and fill space-time).
Let us now consider a simple specific case known as dust (more precise terms used in the literature: pressureless dust, non-interacting incoherent matter, gas of massive, non-interacting particles). This is the model one uses to describe matter in the contemporary universe at scales larger than the one set by the largest gravitational bound states (a few 100 Mpc).
Let us first consider the flat space description (relativistic hydrodynamics). The relevant data are:

1. the four-velocity field: $u = \frac{dx}{d\tau} = u(x)$.
   $u = \gamma(1, \vec{v})$ $(c = 1)$, $\gamma = \frac{dt}{d\tau} = \frac{dx^0}{d\tau}$.
   Comoving frame: $u = (1, \vec{0})$.

2. Energy density measured in the comoving frame: $\rho_0(x)$.

Form a tensor out of ρ_0:

$$T^{\mu\nu} = \rho_0 u^\mu u^\nu \tag{5.60}$$

Then evaluate it in arbitrary inertial frame:

$$T^{00} = \gamma^2 \rho_0 = \rho \tag{5.61}$$

This is the energy density measured in a general inertial frame. Next

$$T^{0i} = T^{i0} = \rho_0 \gamma^2 v^i = \rho v^i \tag{5.62}$$

is the particle flux as seen in a generial inertial frame and

$$T^{ij} = \rho v^i v^j \tag{5.63}$$

is the stress tensor.
Dust satisfies the following equations:

1. The continuity equation (energy conservation)

$$\partial_t \rho + \vec{\nabla}(\rho \vec{v}) = 0 \tag{5.64}$$

2. The Navier-Stokes equation, specialized to vanishing pressure $p = 0$ and vanishing external force $\vec{X} = 0$:

$$\rho \left(\partial_t \vec{v} + (\vec{v}\vec{\nabla})\vec{v} \right) = 0 \tag{5.65}$$

   This expresses momentum conservation.

   For comparison, here is the full Navier Stokes equation:

$$\rho \left(\partial_t \vec{v} + (\vec{v}\vec{\nabla})\vec{v} \right) = -\vec{\nabla} p + \rho \vec{X} \tag{5.66}$$

In relativistic notation, both equations combine into

$$\partial_\nu T^{\mu\nu} = 0 \tag{5.67}$$

($\mu = 0$ is the continuity equation, $\mu = i$ the Navier-Stokes equation.)

Generalization to curved space time: define energy momentum tensor for pressureless dust to be $T^{\mu\nu} = \rho u^\mu u^\nu$, where ρ refers to a local inertial frame (freely falling frame). Energy momentum conservation takes the form $\nabla_\nu T^{\mu\nu} = 0$.

5.1.7 Ideal fluid

A more general continuum model for matter is provided by the ideal fluid. The additional property compared to dust is a non-vanising pressure (field) $p(x)$.

A fluid is called ideal when no dissipative processes occur.

In general, the equation of state takes the form $p = p(\rho, T)$.

For our cosmological purposes, it will be sufficient to consider isentropic fluids, where the equation of state independent of temperature: $p = p(\rho)$.

Dust is the special case $p = 0$.

Ansatz:

$$T^{\mu\nu} = \rho u^\mu u^\nu + pS^{\mu\nu} \tag{5.68}$$

with

$$S^{\mu\nu} = \lambda u^\mu u^\nu + \mu g^{\mu\nu} \tag{5.69}$$

Impose that in the non-relativistic limit the energy-momentum conservation implies the standard hydrodynamik equations (with vanishing external force $\vec{X} = 0$ but non-trivial pressure $p = 0$.)
This implies: $\lambda = \mu = 1$.

Details can be found in: [2], p. 645, [3], p. 35, [4], p. 209, [5], p. 69, [6] chapter 22.

Energy momentum tensor of ideal an fluid:

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + pg^{\mu\nu} \tag{5.70}$$

Remark: An action principle can be found in: [5], p. 69

Definition of ideal fluid makes sense irrespective of the existence of an action principle. Field equations can (and have been) derived without invoking an action principle.

In a comoving frame:

$$T^{\mu\nu} = \mathrm{diag}(\rho, p, p, p) \tag{5.71}$$

This has to be supplemented by specifying the equation of state $p = p(\rho)$. The equations of states relevant for cosmology are of the particularly simple type

$$p = w\rho \ , \quad \text{where} \quad w = \text{const.} \tag{5.72}$$

We have already encountered $w = 0$, which corresponds to dust = non-relativistic particles. This is used whenever the temperature of the universe is so low that the energy distribution is dominated by massive non-relativistic particles. This is called matter dominance. At higher temperature, the rest masses of all (stable) elementary particles can be neglected and all particles are ultrarelativistic. This corresponds to an ideal fluid of (effectively) massless particles and is called radiation dominance. Since photons are massless particles, we expect that the energy momentum tensor of the Maxwell field is of this type.

5.1.8 The Maxwell field as an ideal fluid

Recall that the energy momentum tensor of the Maxwell field is traceless. In a local frame it can be brought to the form

$$T^{\mu\nu} = \mathrm{diag}(\rho, p_1, p_2, p_3) \tag{5.73}$$

with $\rho - p_1 - p_2 - p_3 = 0$. While ρ is the energy density, p_i are the eigenvalues of the stress tensor, i.e., they are pressures. For an isotropic field one has $p = p_i$, and the tracelessness condition gives the equation of state

$$p = \tfrac{1}{3}\rho \; . \tag{5.74}$$

It can be shown generally that $w = \frac{1}{3}$ corresponds to a relativistic fluid of massless particles.

5.1.9 The cosmological constant as an ideal fluid

Bring cosmological constant to the other side of the Einstein equation:

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = \kappa^2 (T_{\mu\nu} + T^{(\Lambda)}_{\mu\nu}) \; , \tag{5.75}$$

where

$$T^{(\Lambda)}_{\mu\nu} = \tfrac{1}{\kappa^2}(-\Lambda) g^{\mu\nu} \tag{5.76}$$

Interpretation in a comoving frame:

$$T^{(\Lambda)}_{\mu\nu} = \tfrac{1}{\kappa^2}\mathrm{diag}(\Lambda, -\Lambda, -\Lambda, -\Lambda) \tag{5.77}$$

Perfect fluid with energy density $\rho = \Lambda/\kappa^2$ and pressure $p = -\Lambda/\kappa^2$. Thus either energy density or pressure is negative.

Remark on terminology: components of the energy-momentum tensor which are not visible (are not sources of electromagnetic radiation) are called dark. Sometimes one uses a narrower meaning (exotic dark matter) which also implies that the dark component only couples gravitational (or through new ultraweak forces) to standard model matter. A dark component of the energy momentum tensor is called dark energy if it is perfectly homogenous. Examples are a cosmological constant, or a homogenous scalar field (quintessence field) with non-vanishing potential energy. In contrast, dark matter refers to stuff which is inhomogenous, at least at sufficiently small scales. Essentially, exotic dark matter means 'elementary particles which are not contained in the standard model and which only couple weakly to it (through gravity or new ultraweek forces). Non-exotic dark matter is non-luminous standard model matter (essentially, everything but stars and hot gas).

5.1.10 The scalar field as an ideal fluid

Recall that the energy momentum tensor for a spatially homogenous field, $\partial_i \phi = 0$, takes the form

$$T_{00} = \frac{1}{2}\dot{\phi}^2 + V(\phi) =: \rho(\phi) \tag{5.78}$$

and

$$T_{ii} = \frac{1}{2}\dot{\phi}^2 - V(\phi) =: p(\phi) \tag{5.79}$$

while $T_{0i} = 0 = T_{ij}$.

The relation between energy density and pressure take the form of the equation of state for an ideal fluid,

$$p(\phi) = w(\phi)\rho(\phi) \tag{5.80}$$

but with a field dependent constant of proportionality, which involves the kinetic and potential energy (density):

$$w(\phi) = \frac{\frac{1}{2}\dot{\phi}^2 - V(\phi)}{\frac{1}{2}\dot{\phi}^2 + V(\phi)} = \frac{T - V}{T + V} \tag{5.81}$$

There is a number of interesting limiting cases

1. Dominance of potential energy, corresponding to a slowly rolling scalar field in a (non-vanishing) potential: $\dot{\phi}^2 \ll V(\phi)$

   $$w = -1 \tag{5.82}$$

   This is equivalent to a cosmological constant.

2. Dominance of kinetic energy, coherent scalar field, $\dot{\phi}^2 \gg V(\phi)$:

   $$w = 1 \tag{5.83}$$

   (Different from standard matter/radiation).

For a scalar potential bounded by zero, $V \geq 0$, the parameter w can take values in the intervall $-1 \leq w \leq 1$. Thus it can also mimic the equations of state for both relativistic and non-relativistic particles. If one allows potential with $V \geq V_0$, with V_0 one can also have $w > 1$. In contrast $w < -1$ is not possible unless one allows the kinetic energy to become negative. (Then one needs to modify the kinetic terms in the scalar action.) Such fields are called phantoms, but it is not presently clear whether such theories make sense. Some motivation for considering phantoms is provided by some observations which indicate that the dominant energy component of the contemporary universe satisfies an equation of state with $w < -1$.

5.1.11 Summary: equations of state

Homogenous and isotropic matter:

$$T_{\mu\nu} = \text{diag}(\rho, p, p, p) \tag{5.84}$$

Isentropic equation of state

$$p = p(\rho) \tag{5.85}$$

indepent of temperature. Specific form in hot big bang cosmology:

$$p = w\rho \tag{5.86}$$

where $w = \text{const}$.

1. $w = 0$
   Matter, better: cold matter, non-relativistic matter, pressureless dust, gas of massive, non-relativistic particles.
   $$p = 0 \tag{5.87}$$

2. $w = \frac{1}{3}$
   Radiation, better: hot matter, relativistic matter, gas of massless or ultra-relativistic massive particles.
   $$p = \frac{1}{3}\rho \tag{5.88}$$
   Equal distribution of kinetic energy on three spatial degrees of freedom. Covers electromagnetic field/photon gas.

3. $w = -1$
   Cosmological constant or scalar field with potential energy dominating kinetic energy.
   $$p = -\rho \tag{5.89}$$

4. $w = 1$
   Homogenous scalar field dominated by kinetic energy.
   $$p = \rho \tag{5.90}$$

5.1.12 Energy conditions

Energy conditions specify assumptions about the behaviour of matter. They are used to prove general statements, like singularity theorems. We only formulate some energy conditions for the specific case of an ideal fluid with energy momentum tensor

$$(T^{\mu\nu}) = \text{diag}(\rho, p, p, p) \tag{5.91}$$

1. The strong energy condition $\rho + 3p \geq 0$.

2. The weak energy condition $\rho + p \geq 0$ and $\rho \geq 0$.
3. The dominant energy condition $-\rho \leq p \leq \rho$ and $\rho \geq 0$.
4. The null energy condition $\rho + p \geq 0$.

Note that that these are not mutually inclusive. E.g., the strong energy condition allows $\rho < 0$ and therefore does not imply whe weak energy condition.

Standard model satisfies all of the above energy conditions. A scalar field only satisfies the null energy condition ($\rho + p = \dot{\phi}^2 \geq 0$). There are 'reasonable' examples of both classical and quantum physical theories which violate all of the above energy conditions. The classical singularity theorems, which, in a sense, proof the existence of an initial singularity, assume the strong energy condition. However the mere observation that the assumed energy condition is not as plausible as thought some time ago does not imply that there is no initial singularity.

5.2 Cosmological solutions

5.2.1 The Friedman equations

Start from Einstein equations:

$$R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G_N T_{\mu\nu} \tag{5.92}$$

Impose that the metric is of FRW form

$$ds^2_{\text{FRW}} = -dt^2 + S(t)^2 \left(\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2_{(2)} \right) \tag{5.93}$$

Then the energy momentum tensor takes the form of an ideal fluid

$$T_{\mu\nu} = (\rho + p) u_\mu u_\nu + p g_{\mu\nu} \tag{5.94}$$

Energy conservation $\nabla^\mu T_{\mu\nu} = 0$ is build in, but provides a short cut to useful equations. To have a well defined problem, we need to specify the equation of state for matter, $p = w\rho$. There are at least three known components: matter ($w = 0$), radiation ($w = \frac{1}{3}$) and (likely) a cosmological constant or other form of dark energy ($w = -1$). (The latter component could also be something more complecited, like a scalar field with field dependent w.)

We will see later that at almost every time one single component dominates the energy condition, while the other can be neglected. Therefore we will first consider an ideal fluid with one single component, but leave the corresponding parameter w arbitrary.

Plugging the FRW ansatz into the Einstein equation gives:

$$\left(\frac{\dot{S}}{S}\right)^2 = \frac{8\pi G_N}{3}\rho - \frac{k}{S^2} + \frac{\Lambda}{3} \tag{5.95}$$

$$\frac{\ddot{S}}{S} + \frac{1}{2}\left(\frac{\dot{S}}{S}\right)^2 = -4\pi G_N p - \frac{k}{2S^2} + \frac{\Lambda}{2} \tag{5.96}$$

Left hand side: geometry. Right hand side: matter, spatial curvature, cosmological constant.
The equations can be reorganized by eliminating $\dot{S}$ from the second equation:

$$\left(\frac{\dot{S}}{S}\right)^2 = \frac{8\pi G_N}{3}\rho - \frac{k}{S^2} + \frac{\Lambda}{3} \tag{5.97}$$

$$\frac{\ddot{S}}{S} = -\frac{4\pi G_N}{3}(\rho + 3p) + \frac{\Lambda}{3} \tag{5.98}$$

The equations are called Friedman equations. Specifically, (5.97) is called the Friedman equation and (5.98) the acceleration equation. Evaluating energy conservation $\nabla^\mu T_{\mu\nu} = 0$ for an idea fluid we get

$$\dot{\rho} + 3\frac{\dot{S}}{S}(\rho + p) = 0 \tag{5.99}$$

S is the scale factor. $\dot{S} > 0$ corresponds to an expanding universe. If $\ddot{S} < 0$ the expansion decelerates, while for $\ddot{S} > 0$ it accelerates. This is also called inflation. For $\Lambda = 0$ one immediately sees that the strong energy condition $\rho + 3p \geq 0$ implies $\ddot{S} \leq 0$, so that inflation is not possible.
We have already seen that a cosmological constant can be interpreted as an ideal fluid with $w = -1$. It is therefore possible to rewrite the terms involving Λ in the Friedman equations as an additional matter term. This will be used frequently in the following.
Since the acceleration equation (5.98) contains second derivatives of the metric it should be considered as the dynamical equation while the Friedman equation (5.97) is a constraint on the initial conditions. Using (5.99) one can show that this interpretation is consistent, because differentiating (5.97) with respect to time does not lead to new independent equations.

In the following we will focus on solving the Friedman equation by making use of energy conservation and the equation of state. It will turn out that this fixes the solution completely, and that the acceleration equation is satisfied automatically. And alternative method is to convert to acceleration equation into a system of autonomous first order equations, or dynamical system, which is then solved. In this approach the Friedman equations is imposed on the initial data. This formulation uses standard techniques from the theory of dynamical systems. In particular one uses fixed point analysis to determine the late time asymptotics of cosmological solutions.

Certain reparametrizations of the Friedman equations are useful. The Hubble function H is defined by the relative growth of the scalar factor:

$$cH := \frac{\dot{S}}{S} \tag{5.100}$$

Here we have reinstalled c, to make explicit that H has dimension lenght^{-1}. (In the literature one often specifies cH in units $\frac{\mathrm{km}}{\mathrm{s}}\frac{1}{\mathrm{Mpc}}$.)
The dimensionless deceleration parameter (or deceleration function) q_0 is defined by

$$q_0 = -\frac{S\ddot{S}}{\dot{S}^2} = -\frac{1}{H^2}\frac{\ddot{S}}{S} \tag{5.101}$$

Due to the minus sign, $H > 0, q_0 > 0$ corresponds to decelerated expansion, while for inflation one has $q_0 < 0$.
The Taylor expansion of $S(t)$ defines an infinite series of dimensionless parameters:

$$S(t) = S(t_0)\left(1 + x - \frac{q_0}{2}x^2 + \frac{q_1}{3!}x^3 - \cdots\right) \tag{5.102}$$

where $x = (t - t_0)H(t_0)$. Explicitly

$$q_n = (-1)^{n+1}\frac{S^{(n+2)}}{H^{n+2}S} \tag{5.103}$$

The density function

$$\Omega := \frac{8\pi G_N}{3H^2c^4}\rho =: \frac{\rho}{\rho_c} \tag{5.104}$$

The critical density

$$\rho_c := \frac{3H^2c^4}{8\pi G_N} \tag{5.105}$$

Value today

$$\rho_{c,0} = 1.8h^2 \cdot 10^{-29}\frac{\mathrm{g}}{\mathrm{cm}^3} \tag{5.106}$$

(where h is the empirical parameter used in measurements of the Hubble function.)
The density function:

$$\chi := \frac{8\pi G_N}{H^2c^4} =: \frac{p}{p_c} \tag{5.107}$$

The critical pressure

$$p_c := \frac{H^2c^4}{8\pi G_N} \tag{5.108}$$

Present value

$$p_{c,0} = 0.16 \cdot 10^{-25}\mathrm{Pascal} \tag{5.109}$$

Rewriting the Friedman equations in terms of H, Ω, q_0, χ:

$$\frac{k}{S^2} - \frac{\Lambda}{3} = H^2(\Omega - 1) \tag{5.110}$$

$$\frac{k}{S^2} - \Lambda = H^2(2q_0 - 1 - \chi) \tag{5.111}$$

Formally we have two algebraic equations for five quantities. While fundamentally we still have two differential equations for three function $S(t), \rho(t), p(t)$ depending on two parameters k, Λ, this algebraic form is useful. For instance, for $\Lambda = 0$ the first equation immediately tells us that $\Omega > 1, = 1, < 1$ or $\rho > \rho_c, = \rho_c, < \rho_c$ implies that $k = 1$ (positive spatial curature), $k = 0$ (vanishing spatial curvature) and $k = -1$ (negative spatial curvature), respectively. Thus we see immediately how the matter content is related to spatial curvature.

5.2.2 Solutions with $k = 0$ and $\Lambda = 0$

For (spatially) flat universes $k = 0$ with vanishing cosmological constant $\Lambda = 0$ and equation of state $p = w\rho$ for the ideal fluid, it is easy to obtain solutions of the Friedman equations in closed form. The case of a cosmological constant (but no matter with an equation of state different from $p = -\rho$) can be covered by reinterpreting the cosmological constant as an ideal fluid. Our universe appears to be spatially flat to a very high degree. Since at a given epoch generically one component in the energy distribution dominates, these simple solutions have physical relevance.

We first use the equation of state $p = w\rho$ to rewrite the energy conservation equation (5.99) as follows:

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{S}}{S} \,. \tag{5.112}$$

Let us first analyse the generic case $w \neq -1$. Integration over time:

$$\int_{\rho_0}^{\rho} \frac{d\rho'}{\rho'} = -3(1+w) \int_{S_0}^{S} \frac{dS'}{S'} \tag{5.113}$$

gives

$$\rho(t) = \rho_0 \left(\frac{S(t)}{S_0} \right)^{-3(1+w)} \tag{5.114}$$

This can be substituted into the Friedman equation (5.97):

$$\left(\frac{\dot{S}}{S} \right)^2 = \frac{8\pi G_N}{3}\rho = \frac{8\pi G_N}{3}\rho_0 \left(\frac{S(t)}{S_0} \right)^{-3(1+w)} \tag{5.115}$$

This differential equation for $\dot{S}$ can be solved by the power law ansatz

$$S = \sigma(t - t_0)^{\beta} \tag{5.116}$$

(Note that S_0, ρ_0 are not $S(t_0), \rho(t_0)$. S_0, ρ_0 are the values of $S(t), \rho(t)$ at lower boundary of the time integral (5.112). In the ansatz we could take $t_0 = 0$, but we prefer not to fix the 'origin' of the time coordinate at this point.)

Using the ansatz we obtain two equations:

1. Matching the powers of t gives

$$\beta = \frac{2}{3(1+w)} \tag{5.117}$$

This fixes the exponent in the time dependence of $S(t)$ of $\rho(t)$ in terms of the parameter of the equation of state, w.

2. Matching the coefficients gives

$$\beta^2\sigma^2 = \frac{8\pi G_N}{3}\rho_0 S_0^{3(1+w)}\sigma^{-1-3w} \tag{5.118}$$

This is one relation between the three constants ρ_0, S_0, σ. It reflects that there are only two independent initial values, one for S one for ρ.

The essential result is

$$S(t) = \sigma(t-t_0)^\beta = \sigma t^{\frac{2}{3(1+w)}} \tag{5.119}$$

which by (5.114) implies

$$\rho(t) = \text{const}(t-t_0)^{-2}\ . \tag{5.120}$$

Note that $\rho(t)$ does not depend on w. In fact, it is immediate to see that any power law for $S \sim t^\beta$ implies that the Hubble function goes like $H \sim t^{-1}$ so that $\rho \simeq H^2 \simeq t^{-2}$.

One can now plugg this solution into the acceleration equation and verify that it is identically satiesfied.

It remains to analyze the special case $w = -1$, which corresponds to a cosmological constant. In this case energy conservation implies

$$\dot\rho = -3\frac{\dot S}{S}(1+w)\rho = 0 \tag{5.121}$$

so that

$$\rho = \rho_0 = \text{const.} \tag{5.122}$$

Thus the energy density is constant. Then the Friedman equation becomes

$$\left(\frac{\dot S}{S}\right)^2 = \frac{8\pi G_N}{3}\rho_0 \tag{5.123}$$

Since the left hand side is non-negative, we conclude $\rho_0 \geq 0$ or equivalently $\Lambda \geq 0$. A negative energy density/cosmological constant is not compatible with flat spatial sections. The Friedman equation can be intergrated elementarily. For an expanding universe, $\dot S > 0$, we obtain

$$S(t) = S_0 e^{\sqrt{\frac{8\pi G_N \rho_0}{3}}(t-t_0)} = S_0 e^{\sqrt{\frac{\Lambda}{3}}(t-t_0)} \tag{5.124}$$

where we used $\Lambda = 8\pi G_N \rho_0 = -8\pi G_N p_0$. Again the acceleration equation is also satisfied.

The solution with $\Lambda > 0$ describes an exponentially expanding universe with constant energy density (and constant negative pressure). This might look like a violation of energy conservation, but here it is useful to note that in a general space-time one cannot define a global quantity 'energy' and then it does not make sense to say that energy is not conserved. The above solution is the de Sitter solution, more precisely the solution (4.219) which covers half of de Sitter space (the static patch).

Let us go through specific cases which illustrate how the behaviour of cosmological solutions depends on the equation of state. We also use this opportunity to introduce a slight generalization. Generically, we find a power law $S \sim (t-t_0)^\beta$, which in general only gives a real S for $t > t_0$. It is natural to extend the solution to $t < t_0$ by $S \sim |t - t_0|^\beta$. The interpretation of the two branches $t > t_0$ and $t < t_0$ will become clear immediately.

Our starting point is the solution for the scale factor

$$S(t) = \begin{cases} \sigma |t - t_0|^\beta & \text{for } w \neq -1 \\ e^{\sqrt{\frac{\Lambda}{3}}(t-t_0)} & \text{for } w = -1 \end{cases} \tag{5.125}$$

where $\beta = \frac{2}{3(1+w)}$.

1. $w > -\frac{1}{3} \Rightarrow 0 < \beta < 1$.
   This includes all forms of standard matter: nonrelativistic matter $w = 0$, radiation $w = \frac{1}{3}$ and also a homogenous scalar field dominated by kinetic energy, $w = 1$.
   One finds that

$$\dot{S} > 0 \,, \quad \ddot{S} < 0 \quad \text{for} \quad t > t_0 \tag{5.126}$$
$$\dot{S} < 0 \,, \quad \ddot{S} < 0 \quad \text{for} \quad t < t_0 \tag{5.127}$$

   Moreover $\dot{S}$ and $\ddot{S}$ diverge for $t \to \infty$. In fact it can be shown that at $t = t_0$ there is a curvature singularity. The fact that the singularity has physical significance is corroborated by observing that the energy density diverges $\rho \sim |t - t_0|^{-2} \to \infty$. (This applies to any power law, not just the range of powers we just considered. Moreover, singularity theorems indicate that such singularities are generic.)

   Due to the singularity, there is no physical transition between the two branches of the solution. The branch $t > t_0$ describes an expanding universe, which starts with an initial singularity at $t = t_0$. The expansion is eternal, but decelerates. The branch $t < t_0$ is obtained by time reflection. It describes a contracting universe, which existed for eternal times and ends in a final singularity (big crunch) at $t = t_0$.

2. $w = -\frac{1}{3} \Rightarrow \beta = 1$.
   Here the time dependence of the scale factor is linear, so that $\ddot{S} = 0$. This is the boundary case between deceleration and acceleration. Again there are two branches, describing an expanding universe starting with a big bang and a contracting universe ending in a big crunch.

   These solutions can be re-interpreted as 'curvature dominance.' When setting $\rho = 0$ and $\Lambda = 0$ in the Friedman equation but allowing non-trivial spatial curvature $k = \pm 1$ we have

   $$\left(\frac{\dot{S}}{S}\right)^2 = -\frac{k}{S^2} \tag{5.128}$$

   Since the left hand side is non-negative we have $k = -1$ and

   $$\dot{S}^2 = 1 \tag{5.129}$$

   Hence, constant negative spatial curvature is formally equivalent to matter with $w = -\frac{1}{3}$.

3. $-1 < w < -\frac{1}{3} \Rightarrow 1 < \beta < \infty$.
   In this range the strong energy condition is violated.
   Now we have

   $$\dot{S} > 0\,, \quad \ddot{S} > 0 \quad \text{for} \quad t > t_0 \tag{5.130}$$
   $$\dot{S} < 0\,, \quad \ddot{S} > 0 \quad \text{for} \quad t < t_0 \tag{5.131}$$

   The branch $t > t_0$ correspondence to accelerated expansion with a power law (power law inflation).

4. $w = -1 \Rightarrow$ Exponential law.
   The case of a cosmological constant.
   This corresponds to accelerated expansion with exponential growth. The solution is eternal, it extends infinitely into past and future without any singularities. (Since general relativity is invariant under time reversal, there is also a time-reflected contracting solution.)

5. $\beta < 0 \Rightarrow w < -1$.
   In this range the null energy condition is violated.
   The law for the scale factor is hyperbolic:

   $$S = \sigma |t - t_0|^{-\alpha}\,, \quad \alpha > 0 \tag{5.132}$$

   The scale factor diverges at $t = t_0$.
   We have

   $$\dot{S} < 0\,, \quad \ddot{S} > 0 \quad \text{for} \quad t > t_0 \tag{5.133}$$
   $$\dot{S} > 0\,, \quad \ddot{S} > 0 \quad \text{for} \quad t < t_0 \tag{5.134}$$

Now the branch $t < t_0$ describes a accelerating expanding universe which existed since infinite time and ends at $t = t_0$ in a final singularity (big rip) when the expansion rate diverges. The branch $t > t_0$ is a contracting universe starting with an initial singularity. The status of these hyperbolic solutions is very unclear. At best they require very exotic physics, such as negative kinetic energy (phantom fields).

5.2.3 Solutions with $\Lambda = 0$ and $k \neq 0$

Matter dominated Friedman solutions

We take the equation of state $p = 0$.

Setting $\Lambda = 0$ in (5.110) and (5.111) gives

$$\frac{k}{S^2} = H^2(2q_0 - 1) = H^2(\Omega - 1) \Rightarrow 2q_0 = \Omega \tag{5.135}$$

Observe that the acceleration is determined by the energy density.

Energy conservation (5.99) then takes the form

$$\frac{d}{dt}(\rho S^3) = 0 \; . \tag{5.136}$$

Since S is the scale factor, this can be interpreted as stating that the energy in a co-expanding volume element is constant. The Friedman equation (5.97) takes the form

$$(\dot{S}^2 + k)S = a \tag{5.137}$$

where

$$a = \frac{8\pi G_N}{3}(\rho S^3) = \text{const.} \tag{5.138}$$

Further manipulating the Friedman equation

$$\pm\frac{dt}{dS} = \frac{\sqrt{S}}{a - kS} \tag{5.139}$$

which can be intergrated

$$\pm(t - t_0) = \int_{S_0}^{S} \frac{\sqrt{S}dS}{\sqrt{a - kS}} \tag{5.140}$$

As a warm-up, take the case with flat spatial sections, $k = 0 \Rightarrow \Omega = 1$. The integral can be performed elementarily, and taking $t_0 = 0$ and $S_0 = 0$ we find

$$\pm\frac{3}{2}\sqrt{a}t = S(t)^{\frac{3}{2}} \tag{5.141}$$

$$\Rightarrow \quad S(t) = \left(\frac{9}{4}a\right)^{\frac{1}{3}} |t|^{\frac{2}{3}} \tag{5.142}$$

This solution is included in the class discussed in the previous section. It is called Einstein-de Sitter cosmos. Summarize data for the expanding branch $t > 0$:

$$S(t) = \left(\frac{9}{4}a\right)^{\frac{1}{3}} t^{\frac{2}{3}} \;, \quad H(t) = \frac{2}{3t} \;, \quad q_0 = \frac{1}{2} \;, \quad \rho = \frac{1}{6\pi G_N t^2} \tag{5.143}$$

Due to the existence of an initial big bang, there is a particle horizon at $r = \frac{3t}{S}$.

A new solution is found for the case of positive curvature, $k = 1 \Rightarrow \Omega > 1$. The Friedman equation

$$(\dot{S}^2 + 1)S = a \tag{5.144}$$

can be integrated to

$$\pm(t - t_0) = \int_{S_0}^{S} \frac{\sqrt{S}dS}{\sqrt{a - S}} \tag{5.145}$$

Setting

$$S = \frac{a}{2}(1 - \cos u) \;, \quad dS = \frac{a}{2}\sin u du \tag{5.146}$$

this becomes

$$\pm(t - t_0) = \int_{S_0}^{S} (1 - \cos u)du \tag{5.147}$$

Thus we get a parametric representation of the curve $S(t)$. With $t_0 = 0$, $S_0 = 0$ we have

$$\pm t = \frac{a}{2}(u - \sin u) \tag{5.148}$$

$$S = \frac{a}{2}(1 - \cos u) \tag{5.149}$$

which is a cycloid. It is immediate that

1. $S \geq 0$.

2. $S = 0$ for $u = 0 \mod 2\pi$, which corresponds to $t = \frac{a}{2}2\pi k$, $k \in \mathbb{Z}$.

3. S has maxima with value $S = a$ for $u = \pi \mod 2\pi$, which corresponds to $t = \frac{a}{2}(\pi + 2\pi k)$, $k \in \mathbb{Z}$.

4. $\frac{dt}{du} > 0$ except at $u = 0 \mod 2\pi$.

5. $S(t)$ is concave, $\ddot{S} < 0$, for $0 \leq t \leq a\pi$. (This is also clear from general arguments: we saw already in our discussion of solutions with $k = 0, \Lambda = 0$ that matter with $w \geq -\frac{1}{3}$ cannot lead to an accelerating scale factor.)

6. $|\dot{S}| \to \infty$ if $S = 0$, indicating singularities for $t = \frac{a}{2}2\pi k$, $k \in \mathbb{Z}$.

The asymptotic form of $S(t)$ for $t = 0$ can be found by expanding

$$t = \frac{a}{2}(u - \sin u) = \frac{a}{12}u^3 + \cdots \quad (5.150)$$

$$S = \frac{a}{2}(1 - \cos u) = \frac{a}{4}u^2 + \cdots \quad (5.151)$$

so that

$$S(t) = \left(\frac{9}{4}a\right)^{\frac{1}{3}} t^{\frac{2}{3}} \quad (5.152)$$

Thus the asymptotics for $t = 0$ is the same as for the flat solution $k = 0$, a big bang singularity. Since the solution becomes singular again at $t = a\pi$, we have found the following result: the solution is formally periodic, but has singularities separating the cycles. Between two singularities, the solution describes a solution which starts with a big bang, expands to a maximal size and then recollapses and end in a singularity (big crunch).

The geometry of space: local vs global

For $k = 1$ the spatial geometry has is a three-sphere. Therefore the volume of space is finite and proportional to the scale factor $S(t)$. At the initial and final singularity one has not only infinite curvature and energy density, but also zero volume of space. Thus initially and finally, space contracts to a point. In contrast, for flat space (and also for negative curvature) the volume is infinite at any time if we work with the simply connected globally Riemannian space determined by the local form of the metric. Solutions with $k = 1, 0, -1$ and therefore sometimes referred to as closed, flat and open universes respectively. There is, however the option to make the volume finite by dividing through the action of a freely acting group of isometries. E.g., flat space can be compactified into a torus.

Open universes

The open universe $k = -1 \Rightarrow \Omega < 1$, can be treated analogous to the closed universe. The Friedman equation

$$(\dot{S}^2 - 1)S = a \quad (5.153)$$

can be integrated to

$$\pm(t - t_0) = \int_{S_0}^{S} \frac{\sqrt{S}dS}{\sqrt{a + S}} \quad (5.154)$$

Setting

$$S = \frac{a}{2}(\cosh u - 1) \ , \quad dS = \frac{a}{2}\sinh u du \quad (5.155)$$

and setting $t_0 = 0$, $S_0 = 0$ one gets the parametric representation

$$\pm t = \frac{a}{2}(\sinh u - u) \tag{5.156}$$

$$S = \frac{a}{2}(\cosh u - 1) \tag{5.157}$$

This solution starts with a big bang at $t = 0$ and grows for infinite time. The asymtotics for $t = 0$ is as for the closed and flat universe $S \sim t^{\frac{2}{3}}$. For late times one finds linear growth $S \sim t$.

Radiation dominated Friedman solutions

No consider radiation dominance, $p = \frac{1}{3}\rho$. This time (5.99) implies

$$\frac{d}{dt}\left(\rho S^4\right) = 0 \tag{5.158}$$

The Friedman equation (5.97) becomes

$$(\dot{S}^2 + k)S^2 = b \tag{5.159}$$

where

$$b = \frac{8\pi G_N}{3}\rho S^4 = \text{const.} \tag{5.160}$$

Integrating

$$\pm\frac{dt}{dS} = \frac{S}{\sqrt{b - kS^2}} \tag{5.161}$$

gives

$$\pm(t - t_0) = \int_{S_0}^{S} \frac{SdS}{\sqrt{b - kS^2}} \tag{5.162}$$

Since $\rho_c = 3p_c$ the equation of state implies $\Omega = \chi$. Using (5.110) and (5.111) this implies

$$\Omega = \chi = q_0 \tag{5.163}$$

Flat radiation dominated Friedman solutions

For $k = 0$ we have (setting $t_0 = 0$ and $S_0 = 0$)

$$\pm t = \frac{1}{2\sqrt{b}}S^2$$

$$\Rightarrow \quad S = \sqrt{2}b^{\frac{1}{4}}|t|^{\frac{1}{2}} \tag{5.164}$$

Summary

$$S = \sqrt{2}b^{\frac{1}{4}}|t|^{\frac{1}{2}}\ , \quad H = \frac{1}{2t}\ , \quad \Omega = \chi = q_0 = 1\ , \quad \rho = \frac{3}{32\pi G_N t^2} \sim S^{-4} \tag{5.165}$$

Closed and open radiation dominated Friedman solutions

For $k = 1$ we have

$$\pm t = \int_0^S \frac{SdS}{\sqrt{b - S^2}} \tag{5.166}$$

Parametric representation of $S(t)$:

$$\begin{aligned} S &= \sqrt{b}\sin u \\ t &= \sqrt{b}(1 - \cos u) \end{aligned} \tag{5.167}$$

(Note that $S_0 = S(t_0) = S(0) = 0$ corresponds to $u = 0$. Thus the lower boundary contributes $-\cos(0) = 1$.) We can find $S(t)$ in closed form

$$S = t^{\frac{1}{2}}\sqrt{2\sqrt{b} - t} \tag{5.168}$$

in the intervall $0 \le u \le 2\pi$ or $0 \le t \le 2\sqrt{b}$. This is a solution which starts pointlike with a big bang, expands to maximal volume, and contracts to a point again.

For $k = -1$ this becomes

$$\pm t = \int_0^S \frac{SdS}{\sqrt{b + S^2}} \tag{5.169}$$

Parametric representation of $S(t)$:

$$\begin{aligned} S &= \sqrt{b}\sinh u \\ t &= \sqrt{b}(\cosh u - 1) \end{aligned} \tag{5.170}$$

or, epxlicitly:

$$S = t^{\frac{1}{2}}\sqrt{2\sqrt{b} + t} \tag{5.171}$$

This starts with a big bang and expands forever.
In all three cases the asymptotics at $t = 0$ is the same $S \sim t^{\frac{1}{2}}$.
The formulae can for radiation dominated Friedman universes can be summarized as

$$S(t) = t^{\frac{1}{2}}\sqrt{2\sqrt{b} - kt} \tag{5.172}$$

$$H(t) = \frac{\sqrt{b} - kt}{t(2\sqrt{b} - kt)} \tag{5.173}$$

$$\Omega(t) = 1 + \frac{k}{(HS)^2} \tag{5.174}$$

$$\rho(t) = \frac{3}{32\pi G_N t^2}\Omega(t)\left(\frac{\sqrt{b} - kt}{\sqrt{b} - \frac{1}{2}kt}\right)^2 \tag{5.175}$$

Friedman solutions with radiation and matter

Consider the eqution where both matter and radiation are present. Energy conservation (5.99) implies

$$\frac{d}{dt}(\rho_R S^4) + S\frac{d}{dt}(\rho_M S^3) = 0 \tag{5.176}$$

where ρ_R and ρ_M are the energy densities of radiation and matter, respectively. Make the ansatz that $\rho_R S^4$ and $\rho_M S^3$ are constant. (Separat energy conservation for radiation and matter.) Then the Friedman equation (5.97) takes the form

$$(\dot{S}^2 + k)S^2 = B + AS \tag{5.177}$$

where

$$A = \frac{8\pi G_N}{3}\rho_M S^3 = \text{const.}\,, \quad B = \frac{8\pi G_N}{3}\rho_R S^4 = \text{const.}\,. \tag{5.178}$$

Thus

$$\pm\frac{dt}{dS} = \frac{S}{\sqrt{B + AS - kS^2}} \tag{5.179}$$

and

$$\pm(t - t_0) = \int_{S_0}^{S} \frac{SdS}{\sqrt{B + AS - kS^2}} \tag{5.180}$$

This integral can be solved explicitly to obtain $t(S)$.

The case $k = 0$ is simplest:

$$\pm(t - t_0) = \left(\frac{2}{3A}S - \frac{4B}{3A^2}\right)\sqrt{B + AS} \tag{5.181}$$

(For $k = \pm 1$ the solution involves inverse trigonometric functions or their hyperbolic analogues.) The ratio between the energy densities is

$$\frac{\rho_R}{\rho_M} \simeq S^{-1} \tag{5.182}$$

(ignoring numerical constants). We see that for $S \to 0$ the radiation density dominates $\rho_R \gg \rho_M$, while for $S \to \infty$ the matter density dominates $\rho_M \gg \rho_R$. Thus one expects that in an expanding universe radiation dominates at early times and matter dominates at late times. This is easy to see for $k = 0$. Expanding the flat solution for $S \to 0$:

$$\pm(t - t_0) = -\frac{4B^{\frac{3}{2}}}{A^2} + \frac{S^2}{3}(3^{-\frac{1}{3}} + B^{-\frac{1}{3}}) + \cdots \tag{5.183}$$

(the contributions $\sim S$ cancel). With $t_0 = \mp\frac{4B^{\frac{3}{2}}}{A^2}$, we have $S \sim t^{\frac{1}{2}}$, which is radiation dominance. For $S \to \infty$ we find

$$t \simeq \frac{2}{3A}S\sqrt{AS} + \cdots \tag{5.184}$$

so that $S \sim t^{\frac{2}{3}}$, which is matter dominance.
For realistic values of parameters the transition between the two asymptotic regimes is fast compared to the growth rate. Therefore it is justified to work with either a radiation dominated or a matter dominated solution instead of the more complicated solution (5.181).

5.2.4 Solutions with $\Lambda \neq 0$ and $p = 0$

We now discuss solutions with cosmological constant (Friedman-Lamaitre solutions), but resctrict ourselves to the case of (non-relativistic) matter. Use ρS^3 is constant, Friedman equation:

$$\dot{S}^2 = \frac{a}{S} - k + \frac{\Lambda}{3} S^2 \tag{5.185}$$

Since $\dot{S}^2$ is non-negative, the equation $\dot{S}{=}0$ defines a separatrix in the (S, Λ) plane, which separates allowed from non-allowed pairs of values. We write the curve in the form

$$\Lambda(S) = \frac{3}{S^2}(k - \frac{a}{S}) \tag{5.186}$$

with $0 \leq S < \infty$. Then allowed values of Λ are those with $\Lambda \geq \Lambda(S)$. The curve $\Lambda(S)$ goes to $-\infty$ for $S \to 0$ and to 0 for $S \to \infty$. It has a critical point, which then must be a maximum, if

$$\frac{d\Lambda}{dS} = 0 \Rightarrow 6kS = 9a \tag{5.187}$$

has a solution. Since $a > 0$, this is the case for $k = 1$, but not for $k = 0, -1$.

Closed universes

For $k = 1$ the curve $\Lambda(S)$ has a maximum at

$$S = S_E = \frac{3a}{2k} = \frac{3a}{2} \; , \quad \Lambda = \Lambda_E = \frac{9a^2}{4k} = \frac{9a^2}{4} > 0 \tag{5.188}$$

Depending on the values of Λ and S, there are six different type of solutions, for which we discuss the asymptotic behaviour and mention some qualitative features.

1. Type M1. $\Lambda > \Lambda_E > 0$. Since this is above the maximum of $\Lambda(S)$, there is no restriction on S. The solution interpolates between matter dominance $S \sim t^{\frac{2}{3}}$ and the exponentially expanding de Sitter solution $S \sim e^{\frac{\Lambda}{3} t}$. There can be a long intermediate period where S is nearly constant (hesitating, loitering).
   Present data suggest that the current universe is in transition between matter dominance and de Sitter behviour, but this transition appears to be rapid, c.f. coincidence problem.

2. Type E. $\Lambda = \Lambda_E$, $S = S_E =$ const. This is a static solution which sits at the maximum of $\Lambda(S)$. It was first proposed by Einstein, is called Einstein cosmos. Note that Hubble constant and redshift vanish for this static solution: $H = q_0 = z = 0$. The solution is unstable in the sense that small perturbation grow exponentially in time. (Dynamical system: saddle point.)

3. Type A1: $\Lambda = \Lambda_E$, $S < S_E$. This starts as a matter dominated universe and approaches the Einstein cosmos at late times.

4. Type A2: $\Lambda = \Lambda_E$, $S > S_E$. This starts as Einstein cosmos and approaches the de Sitter solution at late times. (This also shows that the Einstein cosmos is unstable.)

5. Type M2: $0 < \Lambda < \Lambda_E$ and $S > S_E > 0$. These solutions start on the critical curve with finite S. They approach the de Sitter solution at late time. They have $\ddot{S} > 0$ (acceleration) and $\dot{S}$ has a zero. Thus they shrink starting from finite S to a minimal S and then 'bounce' into expansion, which ultimately becomes de Sitter like.

6. Type O: $\Lambda < \Lambda_E$, $0 < S < S_{\max} < S_E$. These models start from $S = 0$ reach a maximal size and collapse in finite time.

Open and flat universes

Now $\Lambda(S)$ has no maximum. There remain two types of solutions:

1. Type M1: $\Lambda > 0$: infinite expansion starting from $S = 0$.

2. Type O: $\Lambda < 0$: Expansion form $S = 0$ with recollapse in finite time.

5.2.5 Solutions with $\Lambda \neq 0$, matter and radiation

Friedman equation

$$\left(\frac{\dot{S}}{S}\right)^2 = \frac{8\pi G_N}{3}(\rho_M + \rho_R) - \frac{k}{S^2} + \frac{\Lambda}{3} \tag{5.189}$$

Assuming that $\rho_M S^3$ and $\rho_R S^4$ are constant, we obtain

$$\pm(t - t_0) = \int_{S_0}^{S} \frac{SdS}{\sqrt{B + AS - kS^2 + \frac{\Lambda}{3}S^4}} \tag{5.190}$$

This integral can be solved numerically. Asymptotic analysis shows that radiation dominates at early times, while the cosmological term dominates at late times. There might be an intermediate period of matter dominance. See: coincidence problem.

5.2.6 The coincidence problem

To be written.

5.2.7 The relation between distance and redshift

In this section t_0 denotes the present epoch. Recall the relation between redshift and scale factor:

$$1 + z = \frac{S_0}{S} = \frac{\lambda_0}{\lambda} \tag{5.191}$$

where $S_0 = S(t_0)$, $S = S(t)$, etc. Applies to light emitted at time t with wave length λ and observed today with frequency λ_0.
Consider a point source at the coordinates origin, $r = 0$, which generates light of intensity L within the time intervall Δt around time t and in a wave length intervall $\Delta\lambda$. Assuming isotropic expansion, the observed intensity today, L_{obs} is given by

$$L_{obs} F_0 \Delta t_0 \Delta\lambda_0 = L \Delta t \Delta\lambda \tag{5.192}$$

Here F_0 is the geodesic area of the sphere on which the light has been diluted,

$$F_0 = 4\pi S_0^2 r^2 \tag{5.193}$$

We also need to take into account that the wave length and time intervall are both redshifted by a factor $(1+z)$ (both with same factor, due to local constancy of light.) So

$$L = L_{obs}(1+z)^2 4\pi S_0^2 r^2 \tag{5.194}$$

Defining the luminosity distance D_m by

$$D_m = (1+z) S_0 r \tag{5.195}$$

this can be written

$$L = L_{obs} 4\pi D_m^2 \tag{5.196}$$

(which is formally the same relation as in flat space). Since distances cannot be measured directly (and for the most distant objects there are no indirect methods), one would like to have a relation between the distance and the observed redshift. This will depend on cosmological parameters. If alternative methods of measuring distances are available, this can be used to determine cosmological parameters.

The relation can be obtained by combining various Taylor expansion. First one expands $S = S(t)$ in $t - t_0$ and uses the definitions

$$H = \frac{\dot{S}}{S}\,, \quad q_0 = -\frac{\ddot{S}S}{\dot{S}^2} = -\frac{\ddot{S}}{H^2 S}\,, \quad q_n = (-1)^{n+1}\frac{S^{(n+2)}}{H^{n+2}S} \tag{5.197}$$

Then

$$z = -1 + \frac{S_0}{S(t)} = -H_0(t-t_0) + \frac{1}{2}H_0^2(2+q_0)(t-t_0)^2 + \cdots \tag{5.198}$$

Using the variable

$$x = H_0(t_0 - t) \tag{5.199}$$

this becomes

$$z = x(1 + x(1 + \tfrac{1}{2}q_0) + x^2(1 + q_0 + \tfrac{1}{6}q_1) + \cdots) \tag{5.200}$$

(we added the next order from [13]). This can be inverted to give

$$x = z(1 - z(1 + \tfrac{1}{2}q_0) + z^2(1 + q_0 + \tfrac{1}{2}q_0^2 - \tfrac{1}{6}q_1) + \cdots) = H_0(t_0 - t) \tag{5.201}$$

This relation tells how to express the time difference between emission and observation in terms of the redshift. In order to express $D_m = (1+z)S_0 r_0$ in terms of the redshift, it remains to handle r_0. To do so we consider the radial null geodesic of photon emitted at coordinate r at time t and observed at coordinates r_0 today. Recall the modified radial coordinate χ, where $r(\chi) = \sin\chi, \chi, \sinh\chi$ for $k = 1, 0, -1$. We know that this coordinate is related to time by

$$\chi(t) = \chi_0 + \int_{t_0}^{t} \frac{dt}{S(t)} \tag{5.202}$$

(outgoing radial light ray). Then

$$\frac{d\chi}{dt}(t_0) = \frac{1}{S_0}\ , \quad \frac{d^2\chi}{dt^2}(t_0) = -\frac{H_0}{S_0}\ , \quad \frac{d^3\chi}{dt^3}(t_0) = \frac{H_0^2}{S_0}(2 + q_0)\ . \tag{5.203}$$

Taylor expansion of $\chi(t)$:

$$\chi(t) = \chi_0 + \frac{t - t_0}{S_0} - \frac{H_0}{2S_0}(t - t_0)^2 + \frac{H_0^2}{6S_0}(2 + q_0)(t - t_0)^3 + \cdots \tag{5.204}$$

Now we put the source at the origin, $\chi(t) = 0$. Then

$$S_0\chi_0 = (t_0 - t) + \tfrac{1}{2}H_0(t_0 - t)^2 + \tfrac{1}{6}H_0^2(2 + q_0)(t_0 - t)^3 + \cdots \tag{5.205}$$

Then we express $(t_0 - t)H_0$ in terms of z:

$$S_0\chi_0 = \frac{z}{H_0}(1 - \tfrac{1}{2}z(1 + q_0) + \tfrac{1}{6}z^2(2 + 4q_0 + 3q_0^2 - q_1) + \cdots) \tag{5.206}$$

To compute D_m we need $S_0 r_0$ rather then $S_0\chi_0$. Writing

$$D_m = (1+z)S_0\chi_0\frac{r_0}{\chi_0} \tag{5.207}$$

it remains to find $\frac{r_0}{\chi_0}$:

$$\frac{r_0}{\chi_0} = \left\{ \begin{array}{l} \frac{1}{\chi_0} sin\chi_0 \\ 1 \\ \frac{1}{\chi_0}\sinh\chi_0 \end{array} \right\} = 1 - \tfrac{1}{6}k\chi_0^2 + \cdots \tag{5.208}$$

Eliminate χ_0 in terms of z:

$$\chi_0 = \frac{z}{S_0 H_0}\left(1 - \tfrac{1}{2}z(1+q_0) + \cdots\right) \tag{5.209}$$

Putting everything together:

$$D_m = \frac{z}{H_0}\left(1 + \tfrac{1}{2}z(1-q_0) + \tfrac{1}{6}z^2(-1 + q_0 + 3q_0^2 - q_1 - \frac{k}{H_0^2 S_0^2}) + \cdots\right) \tag{5.210}$$

To leading order we recover the Hubble law

$$z = H_0 D_m \tag{5.211}$$

which coincides with the Doppler shift. The curvature $\frac{k}{S_0^2}$ of space only enters at order z^3.
In astronomy the luminosity of stars is historically measured in magnitudes. While the Greeks had six classes of visible stars, it was found in the 19 century that the sensitivity of the eye goes rougly logarithmically with the luminosity (Weber Fechner). The observed apparent magnitude m and the true magnitude M are related by the logarithm of the distance. (M is obtained from m by 'moving' all stars to the same reference distance. The quantitative relation is

$$m - M = 5\log_{10}\frac{D}{10pc} \tag{5.212}$$

Here the reference distance distance in 10 parsec, and the coefficient is determined by fitting the sensitivty of the eye to the logarithm of the luminosity (with a bias to roughly preserve historical data which gives $m \sim 2.5\log_{10} L$). Expressing $D = D_m$ in terms of the redshift one obtains a redshift – brightness relation

$$m = m - 5(1 + \log_{10} H_0) + 5\log_{10} z + \frac{5}{2}\log_{10} e\,(1-q_0)z + \mathcal{O}(z^2) \tag{5.213}$$

5.2.8 The age of the universe

For universes with $S(0) = 0$ (and $\dot{S}$ not identically zero) the age is given by

$$t_0 = \int_0^{S_0} \frac{dS}{\dot{S}} \tag{5.214}$$

(we will consider expanding universes, $\dot{S} > 0$) To find an expression in terms of model parameters and observables, proceed as follows. First solve the Friedman equations for Λ and k:

$$\Lambda = 3H_0^2(\tfrac{1}{2}\Omega_0 + \tfrac{1}{2}\chi_0 - q_0) \tag{5.215}$$

$$-k = H_0^2 S_0^2(-\tfrac{3}{2}\Omega_0 - \tfrac{1}{2}\chi_0 + q_0 - 1) \tag{5.216}$$

Here we already used that these expressions are constant, so that the rhs can be evaluated today. Now use the Friedman equation to solve for $x = \frac{S}{S_0}$:

$$\dot{x} = \sqrt{\frac{8\pi G_N}{3}\rho x^2 - \frac{k}{S_0^2} + \frac{\Lambda}{3}x^2} \tag{5.217}$$

Λ, k can be eliminated using the above equations. To eliminate ρ we use the equation of state. We only treat the cases of matter and radiation. Then

$$\rho(t) = \rho_0 x^{-3(1-w)} \tag{5.218}$$

and we can eliminate ρ_0 in terms of Ω_0

$$\rho_0 = \frac{3H_0^2}{8\pi G_N}\Omega_0 \tag{5.219}$$

Then

$$\dot{x} = H_0\sqrt{\Omega_0 x^{-\delta} - \tfrac{3}{2}\Omega_0 - \tfrac{1}{2}\chi_0 + q_0 + 1 + (\tfrac{1}{2}\Omega_0 + \tfrac{1}{2}\chi_0 - q_0)x^2} \tag{5.220}$$

where $\delta = 1, 2$ for matter and radiation, respectively. Since the equation of state relates pressure and energy density, we can express χ_0 in terms of Ω_0:

$$\chi_0 = (\delta - 1)\Omega_0 \tag{5.221}$$

so that

$$\dot{x} = H_0\sqrt{\Omega_0 x^{-\delta} - \tfrac{\delta+2}{2}\Omega_0 + q_0 + 1 + (\tfrac{\delta}{2}\Omega_0 - q_0)x^2} \tag{5.222}$$

Using x instead of S as integration variable:

$$H_0 t_0 = \int_0^1 dx\,\left(\Omega_0 x^{-\delta} - \tfrac{\delta+2}{2}\Omega_0 + q_0 + 1 + (\tfrac{\delta}{2}\Omega_0 - q_0)x^2\right)^{-\frac{1}{2}} \tag{5.223}$$

If we want to now the age of the universe at a specific redshift z, we take the appropriate upper limit in the integral:

$$H_0 t_z = \int_0^{(1+z)^{-1}} dx\,\left(\Omega_0 x^{-\delta} - \tfrac{\delta+2}{2}\Omega_0 + q_0 + 1 + (\tfrac{\delta}{2}\Omega_0 - q_0)x^2\right)^{-\frac{1}{2}} \tag{5.224}$$

In general, this has to be solved numerically. Let us discuss some cases where the integral can be eliminated analytically.

The age of a matter dominated universe

A special case, where it can be solved analytically is $p = 0$, $\Lambda = 0$ which implies $\delta = 1$, $\chi_0 = 0$ and $2q_0 = \Omega_0$. The integral

$$H_0 t_z = \int_0^{(1+z)^{-1}} dx\,\left(\Omega_0 x^{-1} + 1 - \Omega_0\right)^{-\frac{1}{2}} \tag{5.225}$$

can be found in [25] (or be looked up in [24]). We only display it for the special case $\Omega_0 = 1$:

$$H_0 t_z = \tfrac{2}{3}(1+z)^{-3/2} \tag{5.226}$$

The present age, corresponding to $z = 0$ is therefore $t_0 = \frac{2}{3}H_0^{2/3}$. For $\Omega_0 \neq 1$ we cite two useful results from [25]. The first is for the present age. Since Ω_0, the total matter density is known to be not too much different from 1, it make sense to expand t_0 around $\Omega_0 = 1$. The leading term is

$$H_0 t_0 = \tfrac{2}{3}(1 - \tfrac{1}{5}(\Omega_0 - 1) + \cdots) \tag{5.227}$$

A simplified relation between age and redshift occurs a large redshift (meaning early times), $1 + z \gg \Omega_0^{-1}$. Then

$$H_0 t_z = \tfrac{2}{3}(1+z)^{-3/2}\Omega_0^{-1/2} \tag{5.228}$$

Another useful asymptotic formula applies for small redshifts, $z \ll \Omega_0^{-1}$, i.e., for times close to our present time [13]:

$$(t_0 - t_z)H_0 = \tfrac{2}{3}\Omega_0^{-\frac{1}{2}}(z^{-\frac{3}{2}} + \mathcal{O}(z^{-\frac{5}{2}})) \tag{5.229}$$

The age of a radiation dominated universe

Now ww have $p = \frac{1}{3}\rho$ and $\Lambda = 0$. This implies $\delta = 2$, $\Omega_0 = \chi_0 = q_0$. The integral is

$$H_0 t_z = \int_0^{(1+z)^{-1}} dx \left(\Omega_0 x^{-2} + 1 - \Omega_0\right)^{-\frac{1}{2}} \tag{5.230}$$

The exact solution can be found in [25]. We cite the present age

$$H_0 t_0 = \frac{\sqrt{\Omega_0} - 1}{\Omega_0 - 1} \simeq \tfrac{1}{2}(1 - \tfrac{1}{4}(\Omega_0 - 1) + \cdots \tag{5.231}$$

and the age for large redshift $1 + z \gg \Omega_0^{-1}$:

$$H_0 t_z = \tfrac{1}{2}(1+z)^{-2}\Omega_0^{-1/2} \tag{5.232}$$

The age of a flat universe with matter and cosmological constant

We quote another result from [25]. Take $\Lambda \neq 0$ and matter such that the universe is flat $\Omega_M + \Omega_\Lambda = 1$. Then

$$H_0 t_0 = \tfrac{2}{3}\Omega_\Lambda^{-1/2} \log\left(\frac{1 + \Omega_\Lambda^{1/2}}{\sqrt{1 - \Omega_\Lambda}}\right) \tag{5.233}$$

General properties

Returning to qualitative properties of the general case we observe that the ordner of magnitude of t_0 is set by the Hubble time

$$t_H = \frac{1}{H_0} = \frac{S_0}{\dot{S}_0} \tag{5.234}$$

For a decelerating universe, $\ddot{S} < 0$ the Hubble time is always larger than the actual age, $t_H > t_0$. One way to see it is to consider the line tangent to $S(t)$ at the time t_0. This has the equation

$$y(t) = \dot{S}_0(t - t_0) + S_0 \tag{5.235}$$

The line intersect $S = 0$ at

$$y(t) = 0 \Rightarrow t = t_0 - t_H \tag{5.236}$$

If $\ddot{S} < 0$ then the tangent is above $S(t)$, hence $t < 0$ and therefore $t_H > t_0$.

Chapter 6

Particle physics

6.1 Standard model

6.1.1 Particle content

1. Fermions. All have spin $\frac{1}{2}$.
   (a) Leptons (no strong interactions)
   Charged leptons $Q = -1$.

Particle	e	μ	τ
Mass m	0.511 MeV	105.6 MeV	1777 MeV
τ	$> 4 \cdot 10^{24} y$	$2.2 \cdot 10^{-6}$s	$290 \cdot 10^{-15}$s

   Antiparticle have $\overline{e}, \ldots$ have $Q = 1$, same mass and livetime.

   Neutral leptons = Neutrinos. $Q = 0$.

Paticle	ν_e	ν_μ	ν_τ
Mass m	< 3 eV	< 0.15 MeV	< 18.2 MeV

   The masses are direct search limits. Neutrino oscillations proof that the masses are non-vanishing, with mass differences of order eV. Therefore all masses should be in the eV range. For most purposes, this is effectively massless.

   Antineutrinos $\overline{\nu}_e, \ldots$ are distinct from neutrinos (e.g., reactor neutrinos vs solar neutrinos).

   (b) Quarks (strong interactions)
   Quarks with $Q = \frac{2}{3}$:

Particle	u	c	t
Mass m	1–5 MeV	$1.15 - -1.35$ GeV	174 ± 5 GeV

Quarks with $Q = -\frac{1}{3}$:

Particle	d	s	b
Mass m	3–9 MeV	75–170 MeV	4–4.4 GeV

Masses are not known precisely, because there are no free quarks. The six types of quarks are called quark flavours. Each flavours carries an additional charge with respect to strong interactions. It is called colour and has three distinct values. Quarks which have different colours have the same mass, electric charge etc. If q is a quark field (which has several components, due to the spin. then the colour degree of freedom is represented by adding an additional label:

$$q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} \tag{6.1}$$

where $q = u, d, s, \ldots$ and each q_i is a four-component spinor field.

For each quark there is an antiquark $\overline{u}, \ldots$, which has the opposite charges (including colour) but the same mass. Statistically, quarks which only differ by colour have to be counted as different particles. Thus 6 quarks count as 18 particles, each of which has two spin degrees of freedom Adding the antiquarks, the number of degrees of freedom doubles.

2. Bosons.

(a) Gauge bosons, spin 1.

Particle	γ	$W^{\pm}$	Z^0	G_a
Mass m	0	80.4 GeV	91.19 GeV	0
Charge Q	0	± 1	0	0

There are eight different gluons G_a, which differ in their colour charge.

(b) Higgs boson, spin 0. The only standard model particle which has not been found yet. Search limit for mass: $m > 95.3$ GeV. Charge $Q = 0$.

6.1.2 Interactions

General framework

The typical set-ups for experiments in particle physics are particle scattering and particle decay. Therefore the S-matrix or S-operator is of central importance. The S-operator is the asymptotic time evolution operator, which relates the initial state $|i\rangle$, which has been prepared in the asymptotic past, to the final state $|f\rangle$ which is observed at asymptotically late times. Formally:

$$S = \lim_{t_i \to -\infty, t_f \to \infty} U(t_f, t_i) \tag{6.2}$$

where $U(t_f, t_i)$ is the unitary time evolution operator. The probability to find $|f\rangle$ if the initial state was $|i\rangle$ is

$$W(i \to f) = |\langle f|S|i\rangle|^2 \tag{6.3}$$

The probability must be normalized such that

$$\sum_f W(i \to f) = 1 \tag{6.4}$$

implying that S must be unitary:

$$SS^+ = \mathbb{1} = S^+ S \tag{6.5}$$

In experiments one observes cross sections and decay times. The cross section of a scattering process is

$$\sigma = \frac{R}{\Phi} \tag{6.6}$$

where R is the reaction rate and Φ the particle flux. The dimensions are $[R] = \text{time}^{-1}$ and $[\Phi] = \text{velocity/volume} = 1/(\text{time} \cdot \text{area})$. Therefore the cross section has the dimension of an area. The unit used in particle physics is 1 barn $= 10^{-24}\text{cm}^{-2}$.
For decay processes one either measures the decay time τ or, for short lived particle produced as resonances in scattering processes, the decay width Γ,

$$\tau = \frac{\Gamma}{\hbar} \tag{6.7}$$

A useful formula to relate them is

$$\frac{\tau}{1sec} = \frac{6.58 \cdot 10^{-22}}{\frac{\Gamma}{1\text{MeV}}} \tag{6.8}$$

More detailed information is obtained when measuring differential cross sections $d\sigma$ instead of total cross sections:

$$\sigma = \int_{\text{Phase space}} d\sigma \tag{6.9}$$

A standard example is fixed target scattering, where one measures how many scattered particles have scattering angles (θ, ϕ) (using spherical coordinates centered at the target):

$$\sigma = \int_{\theta,\phi} d\sigma = \int_{\theta,\phi} \frac{d\sigma}{d\Omega} d\Omega \tag{6.10}$$

Let us now sectch how theoretical predictions (S-matrix) and experimental data (cross sections, decay times) are related. We only give examples. Consider first a scattering of two (scalar) particles with momenta k_1, k_2 into N particles with momenta $p_1, \ldots, p_N$. Then the differential cross section is

$$d\sigma = \frac{\prod_i Dp_i}{4\sqrt{(k_1 k_2) - k_1^2 k_2^2}} (2\pi)^4 \delta(4)(k_1 + k_2 - \sum_i p_i) |\langle p_1, \ldots | \mathcal{T} | k_1, k_2 \rangle|^2 \quad (6.11)$$

Here Dp_i are the phase space elements for the outgoing particles

$$Dp_i = \frac{d^3 p_i}{(2\pi)^3 p_i^0} \quad (6.12)$$

the delta functions expresses momentum conservation and $\mathcal{T}$ is related to the S-matrix by

$$S = \mathbb{1} + i\mathcal{T}\delta^{(4)}(\sum_j k_j - \sum_i p_i) \quad (6.13)$$

In going from S to $\mathcal{T}$ one subtracts trivial (non-scattering) processes and splits of momentum conservation.
How then is S (or $\mathcal{T}$) computed for a given theory? The theories underlying the standard model of particle physics are quantum field theories. One starts from a classical field theory, which is defined by a Lagrangian $\mathcal{L}$. The Lagrangian has a free, bilinear part $\mathcal{L}_0$, and an interaction part $\mathcal{L}_I$, which generates the non-linear terms in the equation of motion. As an example let us take

$$\mathcal{L} = -\tfrac{1}{2}\partial_\mu \phi \partial^\mu \phi - \tfrac{1}{2} m^2 \phi^2 - V(\phi) \quad (6.14)$$

where ϕ is a scalar field and $V(\phi)$, the potential, is a polynomial in ϕ. A standard example is

$$V = g\phi^4 \quad (6.15)$$

One standard way to convert this classical field theory into a quantum field theory is to promote the classical fields into operators. Here one should consider the fields $\phi(x)$ as generalized coordinates, analogues to the coordinates x^i of a particle. The quantization of a non-relativistic particle proceeds by postulating

$$[\hat{x}^i, \hat{p}^j] = i\hbar\delta^{ij} \quad (6.16)$$

In field theory the analogue of the momentum p^j is the canonical momentum

$$\Pi(x) = \frac{\partial L}{\partial \dot{\phi}} = \dot{\phi}(x) \quad (6.17)$$

The canonical commutation relations for the scalar field are

$$[\hat{\phi}(x), \hat{\Pi}(y)]_{x^0=y^0} = i\hbar\delta^{(3)}(\vec{x} - \vec{y}) \quad (6.18)$$

Here $\hat{\phi}$ is the field operator associated to the classical field ϕ. Note that the spatial coordinates $\vec{x}$ play the role of a 'continuous index' labelling the generalized coordinate. (Analogous to the index i on x^i). The field operators operate

on the Hilbert space of the theory. This Hilbert space contains states which can be interpreted as describing particles. The mass and spin of these particles are determined by the free part of the Lagrangian.
The formal expression for the S-operator is

$$S = \left(\exp \left[i \int d^4x \, \mathcal{L}_I(\hat{\phi}(x)) \right] \right)_{\text{time ordered}} \tag{6.19}$$

Thus the S-operator is determined by the interaction part of the Lagrangian. Note that since $\hat{\phi}$ are operators with non-trivial commutation relations, one needs to specify an ordering prescription in order that the exponential function (which here is defined by its Taylor series) makes sense. The prescription to be employed here is called time ordering, but we won't explain it here.
One way to compute the S-matrix (or specific scattering amplitudes needed to compute a particular cross section) is perturbation theory. The strength of interactions is controlled by a coupling constant g appearing in $\mathcal{L}_I$ (or by several such constants in more complicated theories) and one can expand in this parameter (provided it is small)

$$S = \sum_{n=0}^{\infty} g^n S^{(n)} \tag{6.20}$$

A convenient way to visualize, organize and perform such calculations is a graphic notation known as Feynman graphs. The Feynmal rules associate to each particle a line, and to each term in $\mathcal{L}_I$ a vertex. The exterior lines of a graph correspond to the initial and final states of the scattering process one wants to compute. Each graph can be interpreted heuristically as a process where particles evolve freely (with the Hamiltonian corresponding to $\mathcal{L}_0$), have interactions corresponding to vertices, propagate again freely, etc. Note that despite this intuitive interpretation, individual graphs of parts of graphs do not correspond to observable proscesses. The scattering probabilities (which are observables) are computed by adding up the contribtions of all Feynman graphs which relate the given initial and final states.

Unitarity bounds

The unitarity of the S-matrix implies bounds on the high energy behaviour. One example known from quantum mechanics is the bound on the cross section of isotropic ('s-wave') scattering

$$\sigma \leq \frac{\pi}{E^2} \tag{6.21}$$

where E is the invariant (center of mass) energy of the process. More generally one has bounds

$$|\mathcal{M}_J| < 1 \tag{6.22}$$

for the partial wave scattering amplitudes (which are defined by expanding the scattering amplitudes $\mathcal{M} = \langle f|S|i\rangle$ in spherical harmonics). This has the

important consequence that theories where the amplitudes and cross sections grow with energy become inconsistent at high energies, because they violate the unitarity bound. Such theories can still make sense, if one interpretes them as (low energy) effective theories which are valid below some energy scale Λ, which is called the cut-off. (This is a physical cut-off, which characterizes a threshold of new physics. In renormalization theory one also introduces cut-offs which serve technical purposes, but are not supposed to have a physical meaning. Such cut-offs have to cancel out ultimately from all observable quantities.)
By dimensional analysis it is easy to see that theories where the coupling has a negative mass dimension potentially have a dangerous high energy behaviour. The reason is that higher powers of the coupling G_X will be accompanied by positive powers of energy. If G_X has dimension mass^{-m} the expected dependence on the coupling and energy schematically is

$$\mathcal{M} \sim G_X + G_X^2 E^m + G_X^3 E^{2m} + \cdots \tag{6.23}$$

This grows with energy and at some energy the unitarity bound will $\mathcal{M}_J \leq 1$ will be violated.
Examples of theories which such couplings are the Fermi theory of weak interactions, ein Einsteins theory of gravity. The coupling constants, the Fermi constant G_F and Newton's constant G_N have dimension mass^{-2} = area. The cut-off energies, where these theories have to be modified by new physics are estimated by

$$G_F \Lambda^2 \simeq 1 \tag{6.24}$$

which is of order of 100 GeV, and

$$G_N \Lambda^2 \simeq 1 \tag{6.25}$$

which is the Planck mass (up to factors 8π) which is of order 10^{16} GeV:

$$\Lambda \simeq \frac{1}{\sqrt{G_N}} \simeq \frac{1}{\kappa} \simeq \frac{1}{L_{\text{Planck}}} \simeq M_{\text{Planck}} \tag{6.26}$$

On the other hand theories with dimensionless couplings are expected to have a good high energy behaviour. Examples are theories where the interactions are mediated by vector bosons, in particular electrodynamics. Here the parameter controlling the perturbation theory is

$$\alpha = \frac{e_0^2}{4\pi\epsilon_0 \hbar c} \simeq \frac{1}{137} \tag{6.27}$$

where e_0 is the elementary charge unit.
A more elaborate theory of vector bosons is used to replace the Fermi theory at higher energies.

Gauge theories

The standard model describes all non-gravitational interactions, namely electromagnetic, weak and strong interactions, through the exchange of vector bosons.

All these interactions are in fact gauge interactions, which can be constructed in terms of symmetry principles. The prototype of a gauge interactions is electrodynamics, which we discuss first.

Electrodynamics

For the benefit of readers not familiar with spinors and the Dirac equation, we will consider a charged, spinless particle for illustration. We will also work in flat-space time for simplicity. A charged spinless particle is described by a complex scalar field with action

$$S = \int d^4x(-\partial_\mu\phi\partial^\mu\bar{\phi} - m^2\phi\bar{\phi} - V(|\phi|)) \tag{6.28}$$

where $\bar{\phi}$ is the complex conjugate of ϕ. The action is invariant under global phase transformations

$$\phi \to e^{ig\chi}\phi \tag{6.29}$$

with parameter χ. It is not invariant under local phase transformations

$$\phi \to \phi' = e^{ig\chi(x)}\phi \tag{6.30}$$

because

$$\partial_\mu(e^{ig\chi(x)}\phi) \neq e^{ig\chi(x)}\partial_\mu\phi \tag{6.31}$$

We can insist on local phase invariance by introducing a suitable covariant derivative D_μ such that

$$D'_\mu\phi' = e^{ig\chi(x)}D_\mu\phi \tag{6.32}$$

This is analogous to the transition from ∂_μ to γ_μ in gravity. Make the ansatz

$$D_\mu = \partial_\mu - igA_\mu \tag{6.33}$$

we find that the connection A_μ must transform as

$$A_\mu \to A'_\mu = A_\mu + \partial_\mu\chi \tag{6.34}$$

This is precisely the transformation behaviour of the vector potential in electrodynamics. Thus the connection can be interpreted as being the Maxwell field. We obtain a field theory with local phase invariance by covariantizing the derivatives and adding the Maxwell action:

$$S = \int d^4x(-D_\mu\phi D^\mu\bar{\phi} - m^2\phi\bar{\phi} - V(|\phi|) - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}) \tag{6.35}$$

Interaction terms can be read off from the terms which are higher then bilinear. Schematically, they are of the form

$$g\phi\phi A\ , \quad g^2\phi\phi AA \tag{6.36}$$

Note that g has the interpretation of coupling constant (electric charge).

In the standard model, the charged particles are fermions with spin $\frac{1}{2}$, which are described by Dirac fields. These are fields ψ with four complex components. The action takes the form

$$S = \int d^4x(\overline{\psi}(i\gamma^\mu D_\mu - m)\psi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}) \tag{6.37}$$

where $\overline{\psi}$ is the Dirac conjugate and γ^μ are the Dirac matrices. This extra matrix structure is related to the spin and will not be discussed here.

Non-abelian gauge theories

Electrodynamics is a gauge theory with the abelian gauge group $U(1)$. Generalizations, where the gauge groups are non-abelian, are relevant for the strong and weak interactions. We discuss this for the (relevant) case where the gauge group is $SU(n)$. Again we take the charged fields to be scalars.
The generalization of abelian gauge theory is then to take an n-dimensional vector of scalar fields

$$\phi(x) = \begin{pmatrix} \phi_1(x) \\ \dots \\ \phi_n(x) \end{pmatrix} \tag{6.38}$$

and to consider transformations

$$\phi(x) \to phi'(x) = U(x)\phi(x) \tag{6.39}$$

where $U(x)$ is a space-time dependent matrix, $U(x) \in SU(n)$. The covariant derivative takes the form

$$D_\mu\phi = (\partial_\mu + gA_\mu)\phi \tag{6.40}$$

where this time the connection A_μ is an n-dimensional matrix, which must transform as

$$A_\mu \to A'_\mu = UA_m uU^+ - \frac{1}{g}\partial_\mu UU^+ \tag{6.41}$$

To write down the analogue of the Maxwell term, one needs the field strength associated with the vector potential A_μ:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + g[A_\mu, A_n u] \tag{6.42}$$

With the commutator, $F_{\mu\nu}$ transforms homogenously

$$F_{\mu\nu} \to F'_{\mu\nu} = UF_{\mu\nu}U^+ \tag{6.43}$$

and Spur($F_{\mu\nu}F^{\mu\nu}$ is invariant.
The locally $SU(n)$ invariant action takes the form

$$S = \int d^4x(-D_\mu\phi^T D^\mu\bar{\phi} - m^2\phi^T\bar{\phi} - V(|\phi|) + \tfrac{1}{2}\text{Spur}(F_{\mu\nu}F^{\mu\nu})) \tag{6.44}$$

For a spinor field it is

$$S = \int d^4x(\overline{\psi}^T(i\gamma^\mu D_\mu - m)\psi + \tfrac{1}{2}\mathrm{Spur}(F_{\mu\nu}F^{\mu\nu})) \tag{6.45}$$

It can be shown that the matrix valued fields $F_{\mu\nu}$ and A_μ take values in the Lie algebra $su(n)$ of the Lie group $SU(n)$. The number of independent fields equals the dimension of the group or algebra and it $n^2 - 1$. Standard vector fields are obtained by expanding the $su(n)$ matrix in a basis T_a=

$$A_m u(x) = \sum_{a=1}^{n^2-1} A_\mu^a(x) T_a \tag{6.46}$$

Thus an $SU(n)$ gauge theory has $n^2 - 1$ massless vector fields A_μ^a.
In a theory with Dirac fermions the interaction vertices have the form

$$g\psi\psi A\ , \quad gAAA\ , \quad g^2AAAA \tag{6.47}$$

The first is a similar to the abelian case and leads to a Coulomb-like interactions. The new feature is the selfcouplings of the A_μ. In other words, the gauge bosons are charged themselves. This leads to a qualitatively different behaviour.
A_μ and $F_{\mu\nu}$ have a geometrical interpretation as the connection and the curvature of an $SU(n)$ principal fibre bundle over space-time. This has similarities with gravity where $G^\rho_{\mu\nu}$ and $R_{\mu\nu\rho\sigma}$ are the connection and the curvature on the tangent bundle. (The associated principle bundle is the bundle of Lorentz frames).
One can consider other Lie groups, simple, semi-simple or abelian. In the abelian case the selfcouplings of the gauge fields are absent. For $G = U(1)$ the Lie algebra is $u(1) = \mathbb{R}$, corresponding to a single real valued gauge field.

Strong interactions and QCD

Strong interactions between quarks are described by an $SU(3)$ gauge theory. Each of the quarks (flavours) has an additional degree of freedom, called colour, which can take three values. More formally a quark transforms in the fundamental representation [3] of $SU(3)$, while antiquarks transform in the antifundamental representation $[\bar{3}]$. According to our general counting, there are 8 vector gauge fields G_μ^a, called gluons. They are themselves charged, more precisely they transform in the adjoint representation [8] of $SU(3)$.
The following description is a useful mnemonic. Call the three quark states red, green and blue, and the three antiquark states antired, antigreen and antiblue, where anti is understood in the sense of additive colour mixing. Then the gluons correspond to all non-trivial changes of colour. In other words they carry both a colour and an anticolour index. There are nine different transitions between three colours, but we have to subtract one particular combination which preserves every colour, leaving eight. (Writing this in matrix form, one obtains an $su(3)$ matrix.)

The qualitative features of the strong interaction between quarks can be illustrated by the static quark–anti-quark potential, which schematically takes the form

$$V \sim \frac{a}{r} + \sigma r \tag{6.48}$$

This is Coulomb-like at short distances, but the energy grows infinitely at large distances. This leads to a permanent binding of quarks into states of zero net colour charge, called confinement. The resulting bound states are elementary particles, called hadrons. While they have no net colour charge, higher multipole moments of the colour force lead to short ranged strong interactions, like those between nucleons and pions.
Leaving exotics aside, there are two classes of hadrons

1. Mesons. These are quark–anti-quark bound states and have integer spin. In terms of the colour terminology, colour and anticolour mix additively to give white. The lightes mesons are the pions, with quark structure

$$\pi^+ = (u, \bar{d}) \ , \quad \pi^0 = (u, \bar{u}) - (d, \bar{d}) \ , \quad \pi^- = (d, \bar{u}) \tag{6.49}$$

   The remaining combination is the η meson

$$\eta = (u, \bar{u}) + (d, \bar{d}) \tag{6.50}$$

   They have spin zero and the following masses, charges and lifetimes

Particle	$\pi^\pm$	π^0	η
Mass m	139.57 MeV	134.17 MeV	547.3 MeV
τ /Γ	$2.6 \cdot 10^{-8} s$	$8 \cdot 10^{-17}$s	1.18 keV
Charge Q	± 1	0	0

2. Baryons. These are bound states of three quarks (or, for anti-baryons of three anti-quarks) and have half-integer spin. In terms of the colour terminology, three basic colours add up to white. The lightest baryons are the nucleons: the proton and the neutron, whith quark contents

$$p = (u, u, d) \ , \quad n = (u, d, d) \tag{6.51}$$

   and masses, charges and lifetimes

Particle	p	n
Mass m	938 MeV	939.6 MeV
τ	$> 10^{33} y$	887s
Charge Q	1	0

There is whole zoo of about 200 hadrons which can all be understood as bound states of quarks. Besides changing the quark content, one can couple the spins

differently and excite angular momentum. Almost all hadrons are very short lived: while states with excited angular momentum or energetically costly spin couplings decay through strong interactions at time scale of 10^{-25}s into other hadrons, the lowest energy states of given quark content can usually still decay through weak or electromagnetic interactions into the leighter leptons. The only long lived hadrons are the nucleons, and the proton is the only possibly stable hadron.
So far we discussed quarks, gluons and hadrons as single particles in the vacuum. Many particle systems with finite temperature and finite pressure can exhibit interesting collective phenomena, which change the behaviour qualitatively, like phase transitions. In QCD there is evidence for a phase transition at $T \approx 150$ MeV. While in the low temperature phase quarks are confined into hadrons, it is believed that above the deconfinement transition quarks and gluons exist as free particles and form a quark-gluon plasma.
For cosmology the following two asymptotic regimes are tractable in equilibrium thermodynamics. At temperature $T \gg 150$ MeV one has a quark-gluon plasma with 18 quark and anti-quark species and 8 gluon species. At temperature $T < 150$MeV quarks (and gluons) are confined into hadrons. At low enough temperature and long enough time scale the only relevant hadrons are the nucleons, since the others decay fast. At time scales longer than several minutes neutrons are only relevant as constitutens of nuclei.
Another relevant concept is baryon number. The standard model interactions preserve this quantum number. A quark (anti-quark) carries $B = \frac{1}{3}$ ($B = -\frac{1}{3}$). A meson carries $B = 0$ and baryon (anti-baryon) carries $B = 1$ ($B = -1$). The matter abundance in the universe today shows that its total baryon number is not zero. This does not seem natural, as matter and antimatter are mirror images of another. It is believed that the initial state of the universe had vanishing baryon number, and that the matter surplus results from processes which violate baryon number conservation. This process is called baryogenesis. Strict baryon number conservation guarantees absolute stability of the proton, as this is the lightest baryon. Some theories which extend the standard model predict new interactions which violate baryon number conservation. In these theories the proton is typically unstable. The fact that it is very long lived provides a strong constraint on such interactions.

Weak interactions: the Fermi theory

Weak interactions manifest themselves in processes involving four fermions. Take for example the β-decay of the neutron:

$$n \to p + e + \overline{\nu}_e \tag{6.52}$$

In the Fermi theory one uses interaction terms (vertices) which are quartic in fermion fields. (We have formulated it here at the level of hadrons, but one could also formulate it with quarks.) The coupling constant appearing at the vertex is the Fermi constant G_F, which has dimension length^2. By dimensional analysis, we expect that the amplitudes behave like

$$\mathcal{M} \sim G_F + G_F E^2 + \cdots \tag{6.53}$$

Unitarity of the S-matrix imposes bounds on the cross section, which are violated at energies of the order

$$E \simeq \Lambda \simeq \frac{1}{\sqrt{G_F}} \tag{6.54}$$

One therefore expects new physics to enter at this scale. The Fermi theory is only a (low energy) effective theory, which is valid below the (physical) cut-off scale Λ.

For theories with massless vector bosons (such as electrodynamics), the coupling constant is dimensionless and the energy dependence of the cross section is compatible with unitarity at high energies. Therefore one might try to reformulate the theory of weak interactions using vector bosons. In terms of graphs the idea is that the four fermion vertex with coupling G_F is replaced by two vertices with dimensionless coupling g_W which couple two fermions to a vector boson. Since weak interactions are short ranged, and since the theory should reduce to Fermi's theory at low energies, the vector bosons must be massive. The schematic relation between Fermi's constant and the new dimensionless coupling is

$$G_F \simeq \frac{g_W^2}{m_W^2} \tag{6.55}$$

where m_W is the mass of the gauge boson. (Heuristically, at low energies the massive vector boson does not propagate and the two fundamental vertices contract into the Fermi four fermion vertex.)

Let us summarize and compare the properties of massive and massless gauge bosons, before proceeding with weak interactions.

Massive vector bosons

The (free part of the) Lagrangian:

$$\mathcal{L} = -\tfrac{1}{4} F_{\mu\nu} F^{\mu\nu} - \tfrac{1}{2} A_\mu A^\mu \tag{6.56}$$

where

$$F_{\mu\nu} := \partial_\mu A_\nu - \partial_\nu A_\mu \tag{6.57}$$

The first term is of Maxwell-type, while the second term is a mass term. Note that this action is **not** invariant under gauge transformations

$$A_\mu \to A_\mu + \partial_\mu \epsilon \tag{6.58}$$

The resulting equation of motion (Proca equation)

$$\partial_\mu(\partial^\mu A^\nu - \partial^\nu A^\mu) - m^2 A^\nu = 0 \tag{6.59}$$

Contracting this equation with ∂^μ we obtain

$$\partial^\mu A_\mu = 0 \tag{6.60}$$

Plugging this back into the original equation we see that the Proca equation is equivalent to the two equations

$$\begin{cases} (\Box - m^2)A_\mu &= 0 \\ \partial^\mu A_\mu &= 0 \end{cases} \tag{6.61}$$

Thus every component of A_μ satisfies the Klein-Gordon equation, the field (and the corresponding particle) has mass m. The second equation imposes one relation between the four components: thus there are only three independent degrees of freedom. To interprete these, consider a particularly simple solution, a plane wave (by Fourier analysis, every solution can be expanded in plane waves):

$$A_\mu = a_\mu e^{ikx} \tag{6.62}$$

The equations of motion translate into

$$\begin{cases} k^2 + m^2 &= 0 \\ k^\mu a_\mu &= 0 \end{cases} \tag{6.63}$$

The first is the dispersion relation/mass shell condition for a free relativistic field/particle with mass m. The second tells us that the polarisation vector a_μ must be orthogonal (in the 4d sense) in the momentum. By a Lorentz transformation we can bring the momentum to the standard form

$$(k^\mu) = (m, 0, 0, 0) \tag{6.64}$$

For a particle this corresponds to the rest frame. In this frame the polarization vector takes the form

$$(a_\mu) = (0, a_1, a_2, a_3) \tag{6.65}$$

Thus the temporal components of the polarization vanishes, and only three polarizations remain. (When Lorentz boosting our frame a bit then one polarization is longitudinal, i.e., parallel to the three momentum $\vec{k}$, while the other two are transversal. This is different to massless gauge bosons, photons, which only have transversal polarization components).

Group theoretical interpretation

Going to the standard form (rest frame) of the momentum vector does not fix our freedom of Lorentz transformations completely. We can still perform rotations. The corresponding subgroup $SO(3) \subset SO(1,3)$ of the Lorentz group characterizes the spin of the field/particle. The four-dimensional vector a_μ of $SO(1,3)$ decomposes into a scalar and a three-dimensional vector with respect to this $SO(3)$. In our (class of) Lorentz frame(s), the scalar is just a_0 and the vector is (a_i), $i = 1, 2, 3$. Thus the second equation of motion eliminates the scalar (or spin 0) component but keeps the vector (or spin 1) part. Of course the choice of the rest frame is a particular choice of coordinates. The decomposition of a Lorentz vector into a spin 0 and spin 1 part can be characterized in a Lorentz-covariant way. The condition which eliminates the spin 0 part is obtained by Lorentz-boosting the condition $a_0 = 0$:

$$k^\mu a_\mu = 0 \tag{6.66}$$

or, in position space

$$\partial^\mu A_\mu = 0 \tag{6.67}$$

One can also identify the condition which eliminates the spin 1 part and keeps the spin 0 part. The covariant version of $a_i = 0$ is

$$k_\mu a_\nu - k_\nu a_\mu = 0 \tag{6.68}$$

or, in position space

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = 0 \tag{6.69}$$

Also note the following: another natural action for a vector field is

$$\mathcal{L} = -\tfrac{1}{2}\partial_\mu A_\nu \partial^\mu A^\nu - \tfrac{1}{2}m^2 A_\mu A^\mu \tag{6.70}$$

Then the equation of motion is just the Klein-Gordon equation for all four compoents,

$$(\Box - m^2)A_\mu = 0 \tag{6.71}$$

The above analysis shows that this lagrangian describes two particles one with spin 1, one with spin 0.

Like massless gauge bosons, massive gauge bosons can be coupled to matter (fermions or scalars). In the static limit, the interaction can be described by a potential, which takes the form

$$V \simeq \frac{e^{-mr}}{r} \tag{6.72}$$

where r is the distance from the matter particle. The Coulomb potential corresponding to massless vector bosons is modified into a Yukawa potential. Due to the exponential term, the effective range of the potential is $\frac{1}{m} < \infty$.

Massless gauge bosons

Dropping the mass term, we get the Maxwell action.

$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{6.73}$$

This lagrangian is invariant under gauge transformations

$$A_\mu \to A_\mu + \partial_\mu \epsilon \tag{6.74}$$

The equation of motion is

$$\partial^\mu(\partial_\mu A_\nu - \partial_\nu A_\mu) = 0 \qquad (6.75)$$

Due to the gauge invariance, the components of A_μ are not independent. We can impose the Lorenz gauge

$$\partial^\mu A_\mu = 0 \qquad (6.76)$$

in which the equation of motion reduces to a massless wave equation

$$\Box A_\mu = 0 \qquad (6.77)$$

The Lorenz gauge eliminates one degree of freedom. In the language of the previous section, it eliminates a spin 0 component and leaves us with spin 1. However, this is not the end of the story. Even in the Lorenz gauge the theory is still invariant under residual gauge transformations with $\Box\epsilon = 0$. Thus there still is one degree of freedom which is unphysical, as it can be arbitrarily changed by gauge transformations. To identify the physical degrees of freedom, consider again a plane wave

$$A_\mu = a_\mu e^{ikx} \qquad (6.78)$$

The equation of motion and Lorenz gauge imply

$$k^2 = 0\,, \quad k^\mu a_\mu = 0 \qquad (6.79)$$

The residual gauge transformation are

$$a_\mu \to a_\mu + \alpha k_\mu \qquad (6.80)$$

where $\alpha \in \mathbb{R}$. Thus one can add an arbitrary vector (anti)parallel to the momentum to the polarization vector without changing the physical state. By a Lorentz transformation we can bring the momentum vector to the standard form (for massless particles there is no rest frame):

$$(k^\mu) = (k^0, \pm k^0, 0, 0) \qquad (6.81)$$

This corresponds to a plane wave travelling along the 1-axis. In this frame, the polarization vector splits into a physical and a gauge part:

$$(a_\mu) = (0, 0, a_2, a_3) + (\alpha, \pm\alpha, 0, 0) \qquad (6.82)$$

The physical polarization components a_2, a_3 are transversal to the direction of the wave. Thus the gauge invariance eliminates the temporal and longitudinal polarization component.

The choice of a standard momentum vector does not fix the freedom of making Lorentz transformations: we can still perform rotations around the axis defined by the (three-)momentum. Under the corresponding $SO(2)$ subgroup of the Lorentz group, the physical polarization components (a_2, a_3) transform as a (two-dimensional) vector. We can go from the transversal to the circular basis for the polarizations by defining

$$a_\pm := a_2 \pm i a_3 \tag{6.83}$$

These transform under $SO(2) \simeq U(1)$ as follows:

$$a_\pm \to e^{\pm i\theta} a_\pm \tag{6.84}$$

Since the group $SO(2) \simeq U(1)$ is abelian, its irreducible representations are one-dimensional. The inequivalent irreducible representations are classified by their helicity $\lambda \in \mathbb{Z}$. The helicity λ representation is

$$a_\lambda \to e^{i\lambda\theta} a_\lambda \tag{6.85}$$

Thus the physical compents of a photon have helicities $\lambda = \pm 1$. For massless particles the concept of helicity replaces the one of spin (which is tied to the full rotation group $SO(3)$). Nevertheless one referes to a particle/field with helicities $\pm\lambda$ as a 'massless spin λ field.' Note that compared to a massive spin 1 field one has lost one degree of freedom. Decomposing the spin 1 representation of $SO(3)$ into representations of $SO(2)$, we obtain $\lambda = 1, 0 - 1$. For a massless vector boson, the helicity zero part becomes a gauge degree of freedom.
One alternative way of dealing with the residual gauge symmetry is to impose the Lorentz gauge and the Coulomb gauge simultanously:

$$\partial^\mu A_\mu = 0 \text{ and } \nabla \cdot \vec{A} = 0 \tag{6.86}$$

or

$$k^\mu a_\mu = 0 \text{ and } \vec{k} \cdot \vec{a} = 0 \tag{6.87}$$

This completely fixes the gauge and leaves us with only the transversal polarizations.
As described in the section on electromagnetic interactions, one can couple massless vector bosons to matter by minimal couling (gauging of global $U(1)$ symmetries). In the static limit photons mediate the Coulomb potential

$$V \simeq \frac{1}{r} \tag{6.88}$$

Back to weak interactions

Now we understand why it makes sense trying to reformulate weak interactions using massive vector bosons. However, if one just adds a mass term by hand, which breaks the gauge invariance explicitly, the problem with unitarity remains. The energy dependence of the cross section still violates unitarity bounds. The problematic processes are those involving longitudinal gauge bosons. It turns out that this problem can by adding scalar fields, so-called Higgs fields to the theory. This enables one to have a mass term for vector bosons while keeping gauge invariance intact. In the following we will discuss this so-called Higgs mechanism. As preparation, we need to talk about a different, but related phenomenon, spontanous symmetry breaking.

Spontanous symmetry breaking

Spontanous symmetry breaking: Consider a complex scalar field $\phi = \phi_1 + i\phi_2$,

$$S = \int d^4x(-\tfrac{1}{2}\partial_\mu\phi\partial^\mu\bar{\phi} - V(|\phi|^2) \tag{6.89}$$

This is invariant under global $U(1) \simeq SO(2)$ transformations. Consider a potential with a degenerate minimum. The standard example is the Mexican hat potential

$$V = \lambda(|\phi|^2 - v^2)^2 \tag{6.90}$$

with $\lambda > 0$ and $v > 0$. Compute extrema

$$\frac{\partial V}{\partial \phi_i} = 4\lambda\phi_i(|\phi|^2 - v^2) \tag{6.91}$$

for $i = 1, 2$. Stationary points are

$$\phi = 0\ , \quad |\phi|^2 = v^2 \tag{6.92}$$

The first is a point in field space, the second a circle $\phi = e^{i\alpha}v \in S^1$. The Hessian is

$$\frac{\partial^2 V}{\partial \phi_i^2} = 8\lambda\phi_i^2 + 4\lambda(|\phi|^2 - v^2) \tag{6.93}$$

$$\frac{\partial^2 V}{\partial \phi_1 \partial \phi_2} = 8\lambda\phi_1\phi_2 \tag{6.94}$$

For $\phi = 0$ the Hessian is negative definite: this is a local maximum. For $|\phi|^2 = v^2$ the Hessian is degenerate, with one positive and one vanishing eigenvalue. To see this observe that the determinante vanishes,

$$\det\left(\frac{\partial \phi}{\partial \phi^2}\right) = 8\lambda(\phi_1^2\phi_2^2 - (\phi_1\phi_2)^2) = 0 \tag{6.95}$$

while the diagonal entries are positive

$$8\lambda\phi_i^2 > 0 \tag{6.96}$$

Using a polar decomponsition, $\phi = e^{i\alpha}\rho$ we see that the potential has local minimum in the radial direction ρ but is flat in the angular direction α, because

$$\frac{\partial V}{\partial \alpha} = 0 \tag{6.97}$$

identically. This shows that there is one massive and one massless real scalar field, the latter is called the Goldstone boson. The appearence of a massless Goldstone boson is a necessary consequence of spontanous symmetry breaking (Goldstone theorem). The Goldstone boson is the modes along the vacuum manifold.

The model exhibits spontanous symmetry breaking: each of the ground states $\phi_\alpha = ve^{i\alpha}$ breaks the $SO(2)$ symmetry of the lagrangian. The set of all ground-states is again $SO(2)$ invariance, and all vacua are mutually related by $SO(2)$ transformations. Mnemonic:

$$\frac{\text{Symmetry of action}}{\text{Symmetry of ground state}} = \text{Set of vacua} \tag{6.98}$$

In the case at hand

$$\frac{SO(2)}{SO(1)} = SO(2) = S^1 \tag{6.99}$$

For n scalar fields with an $SO(n)$ invariant Lagrangian this becomes

$$\frac{SO(n)}{SO(n-1)} = S^{n-1} \tag{6.100}$$

Note that the vacua $e^{i\alpha}v$ are different states. If we want to identify the mass terms in the Lagrangian, we need to single out one ground state, say

$$\phi_0 = v > 0 \tag{6.101}$$

and expand the scalar fields around it (instead of expanding around $\phi = 0$). Parametrize the expansion:

$$\phi_i = \varphi_i + v\delta_{i,1} \tag{6.102}$$

Then the potential becomes

$$V = \lambda((\varphi_1 + v)^2 + \varphi_2^2 - v^2)^2 = \lambda(\varphi_1^2 + 2v\varphi_1 + \varphi_2)^2 \tag{6.103}$$

and the action is

$$S = \int d^4x(-\tfrac{1}{2}\partial_\mu\varphi_1\partial^\mu\varphi_1 - \tfrac{1}{2}\partial_\mu\varphi_2\partial^\mu\varphi_2 - \tfrac{1}{2}m_1^2\varphi_1^2 - 2v\lambda\varphi_1(\varphi_1^2+\varphi_2^2) - \lambda(\varphi_1^2+\varphi_2^2)^2) \tag{6.104}$$

where

$$m_1^2 = 8\lambda v^2 \tag{6.105}$$

is the mass of φ_1, while φ_2 is massless. In addition, we have cubic and quartic interactions.
Before preceeding, let us consider what happens if we do not choose $v^2 > 0$. For $v^2 < 0$, v is imaginary. Set $\mu = iv$ so that the potential becomes

$$V = \lambda(|\phi|^2 + \mu^2)^2 \tag{6.106}$$

This still has a critical point at $\phi = 0$, but the second critical point has disappaered (it would require $|\phi|^2 = -\mu^1 < 0$). The Mexican hat has turned into a curve with minimum at $\phi = 0$. This potential does not show sponanous symmetry breaking. There is a complex scalar field with mass $8\lambda\mu^2$. In the boundary case $v = 0$ the three extrema of the case $v^2 > 0$ have just merged into one single nondegenerate minimum. The Hessian at this point vanishes identically, there is one massless complex scalar.

Let us return to the situation with SSB (spontanous symmetry breaking). A somewhat more elegant procedure is to use an angular decomposition, which is adapted to the structure of the potential. Setting

$$\phi = e^{i\alpha(x)}\rho(x) \tag{6.107}$$

the potential is

$$V = \lambda(\rho^2 - v^2)^2 \tag{6.108}$$

Since V is independent of α it is immediate that there is a S^1 family of degenerate minima for $\rho^2 = v^2$, while $\rho = 0$ is a non-degenerate maximum. Obviously ρ is a massive scalar, while α is the massless Goldstone boson.

To make this explicit, we pick again the vacuum $\phi_0 = v$ and reparametrize the scalar field as:

$$\phi = e^{i\alpha(x)}(\rho(x) + v) \tag{6.109}$$

The potential is

$$V = \lambda((\rho+v)^2 - v^2)^2 = \lambda(2v\rho + \rho^2)^2 = 4v^2\lambda\rho^2 + 4v\lambda\rho^3 + \lambda\rho^4 = \tfrac{1}{2}m_\rho^2 + V_I(\rho) \tag{6.110}$$

where

$$m_\rho^2 = 8v^2\lambda \tag{6.111}$$

For the kinetic term, we need derivatives

$$\partial_\mu\phi = e^{i\alpha}\partial_\mu\rho + ie^{i\alpha}(\rho+v)\partial_\mu\alpha \tag{6.112}$$

The action is

$$S = \int d^4x(-\tfrac{1}{2}\partial_\mu\rho\partial^\mu\rho - \tfrac{1}{2}(\rho+v)^2\partial_\mu\alpha\partial^\mu\alpha - \tfrac{1}{2}m_\rho^2\rho^2 - V_I(\rho)) \tag{6.113}$$

While ρ is massive, α is massless. The terms involving α look like a generalized field dependent kinetic term: this is an example of a sigma model. If we write it out we have a kinetic term for α (with a non-standard normalization[1]) and interaction terms involving the derivatives. α only occurs through its derivative. Therefore the action is invariant under shifts of α. Since α is the angular variable, this is the original $U(1)$ symmetry in disguise. The above action is a (simple) example of a sigma-model. The general form of a sigma model is

$$S = \int d^4x(-\tfrac{1}{2}g_{ij}(\phi)\partial_\mu\phi^i\partial^\mu\phi^j) \tag{6.114}$$

[1] To read of masses from a Lagrangian one needs to separate kinetic and interaction terms, and to bring the kinetic terms to canonical form.

The Higgs mechanism

To discuss the Higgs mechanism we now ‘gauge’ the $U(1)$ symmetry, i.e., we make it a local symmetry by coupling to an abelian gauge field.

$$\mathcal{L} = -\tfrac{1}{2}D_\mu\phi D^\mu\bar{\phi} - V(|\phi|^2) - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{6.115}$$

Local gauge interactions act by

$$\phi(x) \to e^{ig\chi(x)}\phi(x)\ , \quad A_\mu(x) \to A_\mu(x) + \tfrac{1}{g}\partial_\mu\chi(x) \tag{6.116}$$

Observe that by a gauge transformation with $g\chi(x) = -\alpha(x)$ we can eliminate the angular part of ϕ. Thus $\alpha(x)$ can be gauged away, it is not a physical degree of freedom. In particular, the vacua $e^{i\alpha}v$ are related by gauge transformations, and represent the same physical state. The choice $\phi_0 = v$ is a choice of gauge, not the choice of ground state (which is unique). Thus there is no spontanous symmetry breaking. Instead, the gauge boson requires a mass.
To investigate the particle contents, we expand around the minimum $\phi_0 = v$ of V, using the angular decomposition, parametrising $\phi(x) = e^{i\alpha(x)}(\rho(x) + v)$ Working out the term with the covariant derivatives gives

$$D_\mu\phi D^\mu\bar{\phi} = \partial_\mu\rho\partial^\mu\rho + (\rho+v)^2(\partial_\mu\alpha - gA_\mu)(\partial^\mu\alpha - gA^\mu) \tag{6.117}$$

We now use our freedom of making gauge transformation to eliminate α. If we do so we have no freedom of further gauge transformations left: we have fixed the gauge. The scalar covariant derivative terms becomes:

$$D_\mu\phi D^\mu\bar{\phi} = \partial_\mu\rho\partial^\mu\rho + (\rho+v)^2 g^2 A_\mu A^\mu \tag{6.118}$$

The term $v^2g^2A_\mu A^\mu$ is a mass term for the gauge boson, The remaining terms are the kinetic term for ρ and interaction terms. As already mentioned, a mass term of a gauge field is not gauge invariant. However, in the case at hand the local gauge invariance is not broken, but has been fixed by eliminating the field α. In other words, the coupling of a gauge boson to a scalar field allows to give the gauge boson a mass while preserving gauge invariance. This is the Higgs effect (Higgs mechanism).
Thus there are two reason why it is incorrect to call the Higgs effect spontanous symmetry breaking: the ground states is unique and the local gauge symmetry is unbroken. Nevertheless it is standard terminology to say that the Higgs mechanism breaks a gauge symmetry (the $U(1)$ in the case at hand) at a certain scale (set by the mass of the gauge boson). We will follow this practise. A partial justification is provided by the fact that the Higgs mechanism makes the interaction mediated by the gauge boson short-ranged. Thus the symmetry (the gauge boson) does not manifest itself at large scales.
The gauge we have chosen (eliminating α) is called the unitary gauge, because it allows to read off the physical spectrum. The lagrangian takes the following form

$$\begin{aligned}\mathcal{L} \;=\; & -\tfrac{1}{2}\partial_\mu\rho\partial^\mu\rho - \tfrac{1}{2}m_\rho^2\rho^2 - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - \tfrac{1}{2}m_A^2A_\mu A^\mu \\ & -vg^2\rho A_\mu A^\mu - \tfrac{1}{2}g^2\rho^2A_\mu A^\mu - 4\lambda v\rho^2 - \lambda\rho^4\end{aligned} \tag{6.119}$$

with masses

$$m_\rho^2 = 8\lambda v^2 \ , \quad m_A^2 = v^2 g^2 \tag{6.120}$$

We see that the particle content consists of one massive scalar and one massive vector boson. These are $1 + 3 = 4$ degrees of freedom. All these degrees of freedom are physical: there is no gauge degree of freedom left, because it has been used to eliminate one scalar field. The mass terms can be viewed as interaction terms between the constant part of the scalar ϕ, the vacuum expectation value (vev) with the particles which aquire a mass. One can view the presesence of the vev as a modification of the vacuum, which gives a mass to some of the particle. This can be viewed as an (Anti-)Archimedean effect: by swimming in the Higgs vacuum the gauge boson gains mass.

If we consider the gauge invariante lagrangian, the counting of degrees of freedom is different: there is a complex scalar (2 degrees of freedom) and a vector field (4 degrees of freedom). But gauge transformations allow us, as we know from electrodynamics, to eliminate 2 degrees of freedom (-2 degrees of freedom). Thus $2 + 4 - 2 = 0$. The precise particle content depends on the details of the Lagrangian, to be precise of its kinetic, bilinear part. If we consider the same model with $v = 0$, then the kinetic term for the gauge boson is just the standard Maxwell term. In this case the unitary gauge is the combined Lorentz and Coulomb gauge, which eliminates the scalar and the longitudinal mode of the gauge boson and leaves us with two transverse polarizations. But then, we cannot gauge away the scalar α. The spectrum consists off a massless vector boson (2 degrees of freedom) and one complex scalar (2 degrees of freedom). When taking $v \neq 0$, one scalar degree of freedom is converted into the longitudinal model of the vector boson which is now massive. If the scalar lagrangian was not coupled to the gauge field, this would be the Goldstone boson. Therefore the gauge mode is called the would-be Goldstone. 'Higgs effect = the (would-be) Goldstone particle is eaten by the gauge boson'.

Going to a suitable gauge is the easiest way of identifying the vacua (such gauges are usually called 'unitary'). A gauge invariant characterization is provided by correlation functions of gauge invariant objects.

The two different situations $v \neq 0$ and $v = 0$ a referred to as the broken and unbroken phase of the gauge theory, or, better, as the Higgs phase and the symmetric phase. Note that the Higgs vev $\langle\phi\rangle$ does not provide a (gauge-invariant) order parameter, which allows to distinguish the two phases. As we saw all vacua $e^{i\alpha}v$ are gauge equivalent, and by averaging over the gauge orbit we always get zero. A gauge invariant characterization is provided by $\langle|\phi|^2\rangle$, which vanishes in the symmetric phase but is non-vanishing in the Higgs phase.

The Higgs effect in non-abelian gauge theories

The Higgs effect can be generalized to non-abelian gauge theories. In particular, $G = SU(n)$ can be broken by Higgs fields in the fundamental, n-dimensional representation.

Weak interactions as gauge theory

A key observation about weak interaction is that particles can be combined into doublets, whose member differ by one unit of electric charge. Beta decay and related scattering processes

$$\begin{aligned} n &\rightarrow p + e + \overline{\nu}_e \\ n + e_\nu &\rightarrow p + e \end{aligned} \tag{6.121}$$

suggests to combine

$$\begin{pmatrix} p \\ n \end{pmatrix}, \quad \begin{pmatrix} \nu_e \\ e \end{pmatrix} \tag{6.122}$$

Of course the hadrons are composed off quarks, but we can also identify doublets at the quark level

$$\begin{pmatrix} u \\ d \end{pmatrix} \tag{6.123}$$

Introducing massive gauge bosons $W^{\pm}$ with electric charges ± 1, the basic reactions are

$$u \rightarrow d + W^{+} \quad d \rightarrow u + W^{-} \tag{6.124}$$

$$e \rightarrow \nu_e + W^{-}\ , \quad \nu_e \rightarrow e + W^{+} \tag{6.125}$$

etc.
The doublet structure suggest that the group relevant for weak interactions is $SU(2)$. Thus one might try to use an $SU(2)$ gauge theory with Higgs effect and assigns the quarks and leptons to $SU(2)$ doubletts. Since the adjoint representation of $SU(2)$ is three-dimensional, this predicts that there is a third gauge boson, W^3. The corresponding interaction (called neutral current in contrast to the charged currents mediated by $W^{\pm}$) should leave the components of each doublet invariant ($u \rightarrow u$, etc.)
This idea needs to be refined for two reasons:

1. The gauge bosons $W^{\pm}$ are charged with respect to electromagnetic interactions. Therefore weak and electromagnetic interactions are entangled, the combined gauge group cannot be $SU(2) \times U(1)_{em}$.

2. Weak interactions only act on left-handed particles and right-handed anti-particles. Therefore right-handed particles and left-handed anti-particles must sit in singlets. Such a gauge theory is called chiral, because left and righthanded degrees of freedom sit in different multiplets of the gauge group.

The standard model of electroweak interactions, or GSW model (Glashow, Salam, Weinberg) is based on the gauge group $SU(2) \times U(1)_Y$. While $SU(2)$

refers to weak interactions, the group $U(1)_Y$ is not the $U(1)_{em}$ of electromagnetism. The Higgs mechanism works such that $SU(2)\times U(1)_Y$ is broken $U(1)_{em}$, which is identified with a subgroup of $SU(2)\times U(1)_Y$.
Every particle carries two conserved charges: the weak isospin I_3 which refers to the maximal abelian subgroup $U(1)\subset SU(2)$, and the weak hypercharge Y which refers to $U(1)_Y$. The electric charge is identified with $Q = I_3 + Y$. The theory has four gauge bosons: the $SU(2)$ gauge bosons W^+, W^3, W^- are an $SU(2)$-triplet and neutral under $U(1)_Y$. Therefore they carry electric charge $1, 0, -1$. The gauge boson of $U(1)_Y$, denoted B is uncharged under all gauge groups. Since it has the same conserved quantum numbers as W^3, both are allowed to mix. By adding an $SU(2)$-doublet of Higgs fields, one can give a mass to W^+, W^- and

$$Z^0 = \cos\theta_W B + \sin\theta_W W^3 \tag{6.126}$$

while

$$\gamma = -\sin\theta_W B + \cos\theta_W W^3 \tag{6.127}$$

remains massless and is identified with the photon. θ_W is called weak angle or Weinberg angle and parametrizes the mixing between W^3 and B. Since three real degrees of freedom are consumed in giving masses to three gauge bosons, there is one remaining real scalar degree of freedom, the physical Higgs. This corresponds to an observable particle, which has not yet been found (the mass is not predicted).
The Weinberg angle also enters into the relations between couplings,

$$e_0 = g_{SU(2)}\sin\theta_W = g_{U(1)}\cos\theta_W \tag{6.128}$$

and masses

$$m_Z^2 = \frac{m_W^2}{\cos^2\theta_W} \tag{6.129}$$

At energies low compared to the W-mass, one can replace the W-exchange between fermions by an effective four fermion vertex. This gives the relation between the Fermi constant and the couplings of the electroweak theory

$$\frac{G_F}{\sqrt{2}} = \frac{g_{SU(2)}^2}{m_W^2} = \frac{e_0^2}{m_W^2\sin^2\theta_W} \tag{6.130}$$

The Fermi constant G_F and the electric charge e_0 can be measured at low energies. When the W-bosons were discovered $m_W \sim 80$GeV, one could compute the Weinberg angle $\sin\theta_W \simeq 0.23$ and predict the Z-mass $m_Z \simeq 90$GeV, which was then confirmed by experiment.
Quarks and leptons organise into three generations or families, with identical quantum numbers (but different masses). Without loss of generality, we can restrict ourselves to the first family to illustrate their electroweak properties. The lefthanded leptons $(\nu_e, e)_L$ and quarks $(u, d)_L$ sit in $SU(2)$ doublets and therefore have $I_3 = \pm\frac{1}{2}$. By assigning Y-charge $Y = -\frac{1}{2}$ to the lepton doublet

and $Y = \frac{1}{6}$ to the quark doublet we obtain the correct electric charges. The corresponding righthanded degrees of freedom ν_{eR}, e_R, u_R and d_R sit in $SU(2)$ singlets and have $Y = Q = 0, -1, \frac{2}{3}, -\frac{1}{3}$. In order to have a gauge invariant Lagrangian, one cannot introduce mass terms for these fields by hand. Instead, one has Yukawa couplings ($\sim g_{Yuk}\phi\psi\psi$) between fermions and the Higgs doublet. In the unitary gauge the vacuum expectation value of the Higgs field provides the mass terms ($\sim g_{Yuk}\langle\phi\rangle\psi\psi$).

Phase transitions and the Higgs mechanism

The discussion of the Higgs effect gets modified when considering situations with finite temperatures and particle densities. The collective effects of the background thermal bath on gauge bosons, scalars and fermions can be described using a temperature dependent effective potential $V(T)$. As a function of temperature, there might be a phase transition between a Higgs phase with massive gauge bosons at low temperature and a symmetric phase with massless gauge bosons at high temperatures. If fermion masses are generated through Yukawa couplings to the Higgs field, the fermions also become massless.
Such a phase transition is expected to occur for the electroweak interactions at a temperature of about $T = 300\text{GeV}$. At $T \gg 300\text{GeV}$ one has a symmetric, high temperature phase with massless gauge bosons, fermions and sclars. At $T \ll 300\text{GeV}$ one has a broken, low temperature phase with massive W and Z bosons, and with massive quarks and leptons.

6.1.3 Neutrino oscillations and neutrino masses

Above we included right-handed neutrinos in the standard model spectrum. If the neutrinos were massless, it would be consistent to formulate the standard model only with left-handed neutrinos (and right-handed anti-neutrinos). The reason is that (i) right-handed neutrinos do not carry any non-vanishing standard model charges, and (ii) for a massless particle helicity is a conserved quantum number under the restricted Lorentz group (parity is anyway broken maximally by the chiral multiplet structure of electroweak interactions). Once a mass term is present, it mixes left- and right-handed degrees of freedom, and one can only speak of chiral degress of freedom (obtained by projection) but not of indpendent left- and righthanded particles: if a massive particle is left-handed in one Lorentz frame, we can always find a Lorentz boost which overtakes the particle and flips the relative orientation of momentum and spin.
While there are no (undisputed) direct measurements of neutrino masses until today, one has an indirect proof for non-vanishing neutrino masses through the observation of neutrino oscillations. The underlying fact is that there is no symmetry which implies that the interaction eigenstates and mass eigenstates of neutrinos are identical. Interaction eigenstates, $|\nu_a\rangle$, $a = e, \mu, \tau$ are relevant for the production of neutrinos, and for there measurement in detectors. Mass eigenstates $|\nu_j\rangle$, $j = 1, 2, 3$, are relevant for the propagation of neutrinos between source and detector (weak interactions are so weak, that we can safely ignore

interactions on the way between source and detector). Both kinds of states must be related by a unitary matrix $T = (T_{aj})$,

$$|\nu_a\rangle = \sum_j T_{aj}|\nu_j\rangle \tag{6.131}$$

but there is no reason while T must be the unit matrix. Let us explore the consequences of such a mixing matrix. Consider a pointlike static isotropic source of ν_e with energy E located at $\vec{x} = 0$. We put a detector which is sensitive to ν_e at some distance r, and assume that the mass eigenstates propagate with the free relativistic dispersion relation

$$E^2 = \vec{p}_j^2 + m_j^2 \tag{6.132}$$

between source and detector. The propability that a ν_e created at $\vec{x} = 0$ is detected as an ν_e is

$$P(\nu_e \to \nu_e) = |\langle\nu_e|\nu(x)\rangle|^2 \tag{6.133}$$

where

$$|\nu(x)\rangle = \sum_j e^{-iEt+i\vec{p}_j\vec{x}}T_{ej}|\nu_j\rangle \tag{6.134}$$

Using the assumptions listed above one can show that

$$P(\nu_e \to \nu_e) = \sum_i |T_{ei}|^4 + \sum_{i<j} |T_{ei}|^2|T_{ej}|^2 2\cos\frac{2\pi r}{L_ij} \tag{6.135}$$

where $r = |\vec{x}|$ is the distance between source and detector and

$$L_{ij} = \frac{4\pi E}{\Delta m^2_{ij}]};, \quad \Delta m^2_{ij} = |m_i^2 - m_j^2| \tag{6.136}$$

is the oscillation length for oscillations between the i-th and j-th mass eigenstate. Consider for illustration the case of two neutrino species. The mixing matrix can be brought to the form

$$(T_{aj}) = \begin{pmatrix} \cos\theta & \sin\theta e^{i\rho} \\ -\sin\theta e^{-i\rho} & \cos\theta \end{pmatrix} \tag{6.137}$$

Then one can show

$$P(\nu_e \to \nu_e) = 1 - \sin^2 2\theta \sin^2\frac{\pi r}{L} \tag{6.138}$$

Note that in order to have a significant effect, the distance between source and detector may not be small compared to the oscillation length.
In metric units, the oscillation length $L = L_{12}$ is

$$L = 4\pi\frac{E}{\Delta m^2}\frac{\hbar c}{c^4} \simeq 2.5 \text{ Meter } \frac{E/\text{Mev}}{\Delta m^2 c^4/(\text{eV})^2} \tag{6.139}$$

Since direct search limits for the lightes neutrino mass are of order eV, let us consider $\Delta m^2 = 1(\text{eV})^2$. Then the oscillation length for Reaktor neutrinos (1 MeV), meson factories (10 MeV) and accelerator neutrinos (1 GeV) are 2.5 Meter, 25 Meter and 2.5 Kilometer, respectively. To see the effect in the latter case, one has contemplated to send a neutrino beam from Cern to a detector at Gran Sasso.
The above example is not good enough for the comparison with experiments. Here one needs to take into account that there are (at least) three neutrino species which can oscillate. Moreover, interactions can play a significant role when neutrinos travel longer distancs through matter, for example through earth. Here one can have resonance effects: The effective mixing angle in matter θ_M depends on the vacuum angle θ, but also on the electron density N_e. One can have maximal θ_M for a very small θ, if the condition

$$\sqrt{2} G_F N_e = \frac{\Delta m^2}{2E} \cos 2\theta \tag{6.140}$$

is satisfied. (Mikheyev-Smirnov-Wolfenstein or MSW effect).

Overview of experiments

Atmospheric neutrinos.
Reaction of solar wind with upper atmoshpere:

$$\begin{aligned} p, \alpha, \ldots + \text{Air} &\rightarrow \pi^- + X \;, K^- + X \\ \pi^-, K^- &\rightarrow \mu^- + \overline{\nu}_\mu \\ \mu^- &\rightarrow e^- + \overline{\nu}_e + \nu_\mu \end{aligned} \tag{6.141}$$

Ratio of rates of neutrinos:

$$R = \frac{N(\nu_\mu) + N_{\overline{\nu}_\mu}}{N(\nu_e) + N_{\overline{\nu}_e}} \tag{6.142}$$

Naive: $R = 2$. Refined analysis take into account finite live times, detailed kinematics, full list of possible processes. Gives $R = R(E, \phi)$, where E is energy and ϕ the azimut angle. We might expect an interesting ϕ-dependence because neutrinos coming from below have travelled through earth.
Superkamiokande experiment: 50 kT water with 13 000 Cherenkov counters. Sensitive to charged current reactions:

$$\begin{aligned} \nu_e + \text{Matter} &\rightarrow e^+ X \\ \nu_\mu + \text{Matter} &\rightarrow \mu^+ X \end{aligned} \tag{6.143}$$

Without neutrino oscillations they find

$$\frac{R_{exp}}{R_{th}} = 0.68 \pm 0.02 \pm 0.05 \tag{6.144}$$

i.e., a significant deficit in ν_μ. (There also is an interesting ϕ-dependence in the effect). Best fit through oscillations $\nu_\mu \leftrightarrow \nu_\tau$, with $\Delta m_{23}^2 = 3.5 \cdot 10^{-3}(\mathrm{eV})^2$ and $\sin^2 2\theta = 1$.
Generally accepted as proof of neutrino oscillations and hence of neutrino masses (1997). Nobel prize for Kochiba (together with Davis) in 2002.
Solar neutrinos.
Sol produces neutrinos in pp cycle. Balance:

$$4p^+ + 2e^- \rightarrow {}^4He + 2\nu_e + 26.73\mathrm{MeV} \tag{6.145}$$

In detail, six subprocesses of the pp-cycle produce neutrinos, three with continuous, three with discrete spectrum.
Experimental techniques:

1. Chlor detectors (Homestake or Davis experiment, the classic)

$$^{37}Cl + \nu_e \rightarrow {}^{37}Ar + e^- \tag{6.146}$$

   Energy threshold $E > 0.81\mathrm{MeV}$ (only sensitive to solar neutrinos with highest energies).

2. Gallium detectors (Gallex, Sage)

$$^{71}Ga + \nu_e \rightarrow {}^{71}Ge + e^- \tag{6.147}$$

   Energy threshold $E > 0.23\mathrm{MeV}$

3. Water-Cherenkov detectors, (Super-)Kamiokande

$$\nu_e + e^- \rightarrow \nu_e + e^- \tag{6.148}$$

   no threshold, but background.

Solar neutrino unit:

$$1\mathrm{SNU} = \frac{1\mathrm{Einfang}}{10^{36}\mathrm{Targetatome}\ \cdot\ \mathrm{Sekunde}} \tag{6.149}$$

Experiments showed, from the beginning $\simeq 2$ SNU, while theory predicted $\simeq 6$ SNU. Solar neutrino problem. (For a long time, solar physics was not known well enough to be excluded as an alternative explanation. Today it is accepted that the solar neutrino problem is solved by neutrino oscillations.)
SNO experiment (Sudbury Neutrino Observatory) June 2001, salt(?) mine in Canada, 2 km deep, using 1000 tons of heavy water D_2O, borrowed from nuclar reactors. Cherenkov detectors. Measurements:

1. Inelastic processes from charged currents. Example:

$$\nu_e + d \rightarrow u + e^- \tag{6.150}$$

   which translates to

$$\begin{aligned} \nu_e + n &\rightarrow p + e^- \\ \nu_e + D &\rightarrow p + p + e^- \end{aligned} \tag{6.151}$$

2. Elastic processes from charged and neutral currents. Example

$$\nu_X + e^- \to \nu_X + e^- \tag{6.152}$$

which might be mediated by either W or Z boson.

3. Inelastic processed generated from neutral currents.

$$\begin{aligned} \nu_X + q &\to \nu_X + q \\ \nu_X + D &\to p + n + \nu_X \end{aligned} \tag{6.153}$$

The charged current events from ν_e are compatible with those observed by Superkamiokande. But one also measures elastic processes induces by ν_μ and ν_τ and can sum up to obtain the total neutrino flux, which is consistent with the solar standard model.
Combining all neutrino data (including cosmological data, i.p. WMAP) one finds summed neutrino masses

$$\sum_i m_i < 0.71\text{eV} \tag{6.154}$$

mass differences

$$\Delta m^2 \simeq 10^{-3}(\text{ev})^2 \tag{6.155}$$

and large mixing angles.
From the total decay width of the Z one can conclude that there are precisely 3 standard model families with light neutrinos ($2m_\mu < m_Z$).

6.2 Beyond the standard model

6.2.1 Grand unified (gauge) theories

Neutrino data give information about possible physics beyond the standard model. The extemely small masses (and large angles) need explanation. The quantum numbers also give a hint. With non-vanishing masses one needs to include the ν_R and 16 particles per family (counted quarks weighted with colour). The quantum numbers are such that these fit precisely into the 16-representation of the group $SO(10)$.
One idea of extending the standard model is that there is a single (simple) gauge group, which is reduced to the standard model at a high scale by the Higgs mechanism:

$$SU(3) \times SU(2) \times U(1) \subset G_{GUT} \tag{6.156}$$

Such a GUT theory (grand unified theory) has one single coupling constant and should predict the standard model couplings. The above observation indicates that $SO(10)$ is an interesting candidate for G_{GUT}. If for example one takes $SU(5)$ (which still is large enough), then each family fills three representations $10 + \overline{5} + 1$. Moreover $SU(5)$ predicts proton decay at a rate incompatible with experiment. Another good feature of $SO(10)$ GUTs is that by the so-called see-saw mechanism one can naturally obtain very small neutrino masses.

6.2.2 Gravity and Superunification

Einstein's theory of gravity can be reformulated as a 'gauge theory for massless spin 2 particles.' In this reformulations similarities with the other interactions become visible, one can try to quantize gravity in a perturbative framework.
The key element is linearization. Define a gravity field by taking the difference of the Riemannian metric $g_{\mu\nu}$ and a reference metric which satisfies the vacuum Einstein equations. In absence of a cosmological constant, Minkowski space is a solution, and it is natural to consider the deviation from Minkowski space as the gravitational field (as long as these deviations are small).

$$\psi_{\mu\nu} = g_{\mu\nu} - \eta\mu\nu \qquad (6.157)$$

(We have set $\kappa = 1$.) Plugging this into the vacuum Einstein equation

$$R_{\mu\nu} = 0 \qquad (6.158)$$

one obtains a wave equation (with respect to the flat metric) plus non-linear terms

$$\Box_\eta \psi_{\mu\nu} = 0 + \text{non-linear terms in } \psi_{\mu\nu} \qquad (6.159)$$

To be precise, this holds in a particular gauge. The diffeomorphism invariance of the full non-linear theory

$$x^\mu \to f^\mu(x) \qquad (6.160)$$

translates into a gauge invariance

$$\psi_{\mu\nu} \to \psi_{\mu\nu} + \partial_\mu \epsilon_\nu + \partial_\nu \epsilon_\mu \qquad (6.161)$$

Note that the gauge parameters form a vector field ϵ_μ. To obtain a standard wave equation one needs to impose the Einstein gauge (also called Weyl gauge):

$$\chi_\mu := \partial_\nu \psi^\nu_\mu - \frac{1}{2}\partial_\mu \psi^\nu_\nu \stackrel{!}{=} 0 \qquad (6.162)$$

This is analogous to the Lorenz gauge in electrodynamics.
The symmetric tensor field $\psi_{\mu\nu}$ has ten independent coordinates, but the four gauge conditions make four of them dependent on the others. Thus there remain six degerees of freedom. However, the Einstein gauge does not fix the gauge freedom completely. One can still perform gauge transformations with the gauge parameters restricted by $\Box\epsilon_\mu = 0$. Thus four degrees of freedom can be changed at will and are gauge degrees of freedom. The number of physical degrees of freedom is $10 - 4 - 4 = 2$. One could eliminate the four remaining degrees of freedom by a further gauge condition, analogous to the Coulomb gauge, at the expense of Lorentz invariance.
To identify the mass and spin of the field $\psi_{\mu\nu}$ (which become, by quantization, the mass and spin of the associated particle, the graviton) we consider a plane wave solution:

$$\psi_{\mu\nu} = t_{\mu\nu} e^{ikx} \qquad (6.163)$$

The wave equation implies $k \cdot k = 0$ which show that we the field/particle is massless. By a Lorentz transformation, the momentum vector can be brought to the standard form

$$k = (k^0, \pm k^0, 0, 0) \tag{6.164}$$

which corresponds to a plane wave along the 1-axis. The polarization tensor $t_{\mu\nu}$ is symmetric $t_{\mu\nu} = t_{\nu\mu}$ and subject to the gauge condition

$$k_\nu t_\mu^\nu - \frac{1}{2} k_\mu t_\nu^\nu = 0 \tag{6.165}$$

The residual gauge transformation are

$$t_{\mu\nu} \to t_{\mu\nu} + k_\mu \tilde{\epsilon}_\nu + k_\nu \tilde{\epsilon}_\mu \tag{6.166}$$

It can be shown that the polarization tensor takes the form

$$(t_{\mu\nu}) = \begin{pmatrix} 0_2 & 0_2 \\ 0_2 & t_{ab} \end{pmatrix} + \text{gauge degress of freedom} \tag{6.167}$$

where $a, b = 2, 3$, and

$$(t_{ab}) = \begin{pmatrix} t_{22} & t_{23} \\ t_{32} & t_{33} \end{pmatrix} = \begin{pmatrix} t_{22} & t_{23} \\ t_{23} & -t_{22} \end{pmatrix} \tag{6.168}$$

Observe that t_{ab} transforms as a second rank symmetric traceless tensor under the subgroup $SO(2)$ of the Lorentz group which consists off rotations around the 1-axis (in a frame of reference where the momentum vector takes the above standard form). This shows that the massless particle has spin 2, or more correctly, helicities $h = \pm 2$. We can go to a 'circular' basis of polarizations by defining

$$t_\pm = t_{22} \pm i t_{23} \tag{6.169}$$

These components transform as

$$t_\pm \to e^{\pi 2 i \theta} t_\pm \tag{6.170}$$

under the $SO(2)$, which shows explicitly that the helicities are ± 2. In summary the linear part of Einsteins theory tells us that the dynamical degrees of freedom correspond to a massless field/particle with helicities ± 2.

Group theoretical aspects

$\psi_{\mu\nu}$ is a symmetric tensor field. This realizes a reducible representation of the Lorentz group, because the trace is a one-dimensional irreducible representation (Lorentz scalar). The irreducible Lorentz representations are the symmetric traceless part (nine-dimensional) and the trace (Lorentz scalar, one-dimensional):

$$\begin{aligned} \psi_{\mu\nu} &= (\psi_{\mu\nu} - \tfrac{1}{4}\eta_{\mu\nu}\eta^{\rho\sigma}\psi_{\rho\sigma}) + \tfrac{1}{4}\eta_{\mu\nu}\eta^{\rho\sigma}\psi_{\rho\sigma} \\ &= \phi_{\mu\nu} + \phi\eta_{\mu\nu} \end{aligned} \tag{6.171}$$

We can identify the spin content by decomposing these representations with respect to an $SO(3)$ subgroup. We can do it by splitting Lorentz indices into a time index and three space indices. The spin 2 representation is the symmetric traceless tensor of $S0(3)$. It is five-dimensional (corresonding to spin projections $+2, +1, 0, -1, -2$ onto a distinguished axis). The spatial-spatial components ψ_{ij} of $\psi_{\mu\nu}$ give a spin 2 representation together with a spin 0 representation (the trace):

$$\psi_{ij} = (\psi_{ij} - \tfrac{1}{3}\delta_{ij}\delta^{lk}\psi_{lk}) + \tfrac{1}{3}\delta_{ij}\delta^{lk}\psi_{lk} \tag{6.172}$$

The temporal-temporal component ψ_{00} is a another scalar. The temporal-spatial components ψ_{0i} provide an $SO(3)$ vector. In total we have one spin 2, one vector (spin 1) and two scalars (spin 0):

$$10 = 5 + 3 + 1 + 1 \tag{6.173}$$

Like in the case of vector bosons, one can project out the lower spin components in a covariant way. We already saw above that the by splitting of the trace we can eliminate one Lorentz scalar,

$$10 = 9 + 1 \tag{6.174}$$

For a field we can further impose

$$\partial^\mu \partial^\nu \phi_{\mu\nu} = 0 \tag{6.175}$$

This removes another scalar. Imposing the stronger condition

$$\partial^\mu \phi_{\mu\nu} = 0 \tag{6.176}$$

which eliminates one Lorentz vector, corresponding to one spin 1 and one spin 0 field. Thus we are precisely left with the spin 2 part.

The Einstein gauge also removes a Lorentz vector, i.e, one spin 1 and one spin 0 component. We are left with six degrees of freedom, corresponding to spin 2 plus spin 0. The residual gauge symmetry corresponds to a Lorentz vector and hence to a spin 1 and a spin 0 representation. What is then left?

Since the graviton is massless, the relevant concept for characterizing the physical degrees of freedom is not spin (related to an $SO(3)$ subgroup) but helicity (related to an $SO(2)$ subgroup). The $SO(2)$ content of $\psi_{\mu\nu}$ is obvious:

$$10 = 1 \cdot [\lambda = \pm 2] + 2 \cdot [\lambda = \pm 1] + 4 \cdot [\lambda = 0] \tag{6.177}$$

As we saw above the $[\lambda = \pm 2]$ is physical, while the other helicities are gauge degrees of freedom.

Interacting massless spin 2 fields and the quantization of gravity

On can then write down a free (bilinear) gauge invariant action for the field $\psi_{\mu\nu}$. The non-linear terms, which we discarded so far correspond to an infinite series of vertices which describe three-graviton interactions, four-graviton interactions

and so forth. (Since the series does not terminate, sucah theory is called non-polynomial.) When starting with Einstein gravity, the coefficients of all these vertices are fixed. It is natural to ask whether one can find other interactions for a massless spin 2 field. If one insists on gauge invariance, the answer is no: Einstein gravity is the only known consistent interaction for massless spin 2 fields. Once one starts with a three-graviton vertex, gauge invariance forces one to introduce all the higher vertices, with prescribed coefficients and the infinite sum gives the full non-linear Einstein-Hilbert action. One also observes that the flat Minkowski metric used in defining the field theory combines with the graviton field into a Riemannian metric, while the gauge symmetry translates into diffeomorphism invariance. This uniqueness of interactions for massless spin 2 fields distinguishes gravity from other interactions.
The coupling constant controlling interactions of gravitons is $\kappa \simeq \sqrt{G_N}$. Since G_N has the dimension length^2 (in units $\hbar = c = 1$), the amplitude behaves like

$$\mathcal{M} \sim\sim G_N + G_N^2 E^2 + \cdots \tag{6.178}$$

and we expect problems with unitarity bounds at energies

$$\Lambda \simeq \frac{1}{\sqrt{G_N}} \simeq \frac{1}{\kappa} \simeq \frac{1}{L_{\text{Planck}}} \simeq M_{\text{Planck}} \tag{6.179}$$

Like in the case of Fermi's theory of weak interactions, one can now speculate that there is a substructure which resolves the gravitational vertices at energies above this scale. There is one particular candidate, string theory, which replaces point-like elementary particles by one-dimensional objects called strings. Elementary particles are identified with specific vibration modes of strings. Every consistent quantum theory of strings necessarily contains a graviton, and it has been shown that the high energy behaviour is such that one obtains a unitary theory. The basic effect is that at high energies higher vibration modes of strings are excited, resulting in a 'softer' high energy behaviour. (The interaction region 'spreads out'.)
Another interesting feature of string theory is that it has one single coupling constant, g_{string}, while it does not only describe gravity, but also gauge interactions and matter. In contrast to unified theories, where there is only one gauge coupling, but gravity remains independent, it is a superunified theory, where all interactions are controlled by one single coupling constant. These appealing features come with a heavy prize, as consistent string theories require a lot of structure which have not yet been observed in our universe: additional space dimensions and various particle beyond the standard model. The most severe problem is that so far one can only make testable predictions when making various additional assumptions (which is then not so much different from phenomenological model building).
For completeness let us mention that there are also other attempts to formulate a quantum theory of gravity. All have in common that they are not based on perturbation theory and Feynman graphs, but attempt to quantize the full nonlinear theory without singling out a reference background around which one

expands ('background independence'). Typically, a unification of matter and gravity (in the sense of superunified theories) is not attempted. The most active direction is 'loop quantum gravity', which is based on the Hamiltonian formulation of Einstein's theory but uses, instead of the metric, one-dimensional loops as the basic variables. While one can develop a kind of quantum geometry of space-time along this line, it appears to be difficult to get back to a semi-classical space-time. Advantages and disadvantages are thus roughly complementary to the perturbative approach used in string theory.

6.3 Remarks on the literature

Group theoretical aspects of wave equations are discussed in [10]. Gravity as spin 2 field theory is discussed in [11].

Chapter 7

Thermodynamics

Remark:
The geometrical scale factor is renamed $R(t)$, while S is reserved for the entropy.

7.1 Overview

7.1.1 Conceptual remarks

Matter in the universe is a many-particle system which should be desribed by (quantum) statistical physics. Macroscopic (thermodynamic) quantities like temperature, pressure, particle density etc. can be computed once the underlying probability distributions for the relevant microscopic quantities are known (example: particle momenta $\rightarrow$ pressure).

Thermodynamic equilibrium is a distinguished behaviour which is particularly simple and, according to experience, occurs frequently. Thermodynamic equilibrium is a stationary and maximally homogenous state, which can be characterized as a maximum of the (Shannon) entropy (average missing information about the microstate, if a full set of independent macroscopic quantities is specified). From the entropy principle (second law of thermodynamics) one can obtain the equilibrium (probability) distributions.
In principle this principle should be derivable from microscopic physics, starting from the Liouville theorem (conservation of phase space volume) in its quantum field theoretic version (which incorporates particle production). A frequently used approximation is the Boltzmann equation, which is an integro-differential equation characterizing the modification of distributions by interactions (collision term). So far, a derivation of the entropy principle has only be obtained for sufficiently simple systems (dilute gas, ...). But the Boltzmann equation and related methods have the advantage that they can also be applied to non-equilibrium systems.

In the treatment of thermodynamics of the expanding universe we will assume (and claim) that the universe is, 'most of time' in an (approximate) state of thermodynamic equilibrium. Specifically, the expansion is assumed to be reversible, quasistatic and adiabatic (i.e., entropy is conserved). Clearly, this can only be an approximation. A FRW solution, which is only the coarsest approximation to our universe, has no time-like Killing vector and is not stationary. The 'real' world around us is full of irreversible processes. The formation of structures out of baryonic matter, from stars to galaxies and superclusters, to planets, live-forms and ourselves, is an irreversible, non-equilibrium process. Nevertheless, it is plausible to treat the evolution of the universe as a sequence of approximate equilibrium states, because the structured baryonic matter is immersed into the CMB, which is a photon gas with a perfect equilibrium distribution with temperature of 2.7 K. The particle density and entropy of the universe is dominated by the CMB, while the baryonic matter is only a small fraction. This illustrate that (i) it makes sense to analyse the universe as an equilibrium system and that (ii) this picture has limitations because many interesting structures in the universe are non-equilibrium systems. Apart from baryonic matter, there may be other 'relicts', resulting from particles which decouple from the equilibrium, like: massive neutrinos and other massive particles (non-standard model particles, magnetic monopoles, cold dark matter). There might also be extended structures (cosomic strings, domain walls) which result from non-equilibrium processes.

We will see in the following that equilibrium thermodynamics gives a consistent description of many interesting phenomena. But the reader should be aware that this treatment is far from complete. A full treatment requires to justify the applicability of thermodynamic equilibrium (beyond mere self consistency), and where necessary, the use of non-equilibrium methods. This usually means to use the Boltzmann equation, and the analysis becomes rather involved.

In the following we will give a preliminary discussion of the relevant concepts. For the orientation of the reader, we also give a brief overview of the thermal history of the universe. Many statements are given without proof. Some but not all of these statements will be justified later in the chapter.

7.1.2 Temperature, energy, expansion

The FRW line element:

$$ds^2 = -dt^2 + R(t)^2 ds^2_{(3)} \tag{7.1}$$

Note that the scale factor has been renamed, $S(t) \to R(t)$. In this chapter, S denotes the entropy.

We already state here (without) proof that the relation between temperature $T(t)$ and scale factor $R(t)$ is

$$\frac{T}{T_0} = \frac{R_0}{R} \tag{7.2}$$

for radiation dominance and

$$\frac{T}{T_0} = \frac{R_0^2}{R^2} \tag{7.3}$$

for matter dominance. (Here $T = T(t)$, $T_0 = T(t_0)$, etc.) Thus temperature scale as $T \propto R^{-1}$ and $T \propto R^{-2}$, respectively.
In thermodynamic equilibrium, temperature characterizes the average energy per degree of freedom. More precisely each degree of freedom carries the average energy $\frac{1}{2}k_B T$, where

$$k_B = 1.38 \cdot 10^{-23}\frac{\text{J}}{\text{K}} = 8.62 \cdot 10^{-5}\frac{\text{eV}}{\text{K}} \tag{7.4}$$

is the Boltzmann constant. This implies that we can measure temperature in energy units. Setting $k_B = 1$ on top of $\hbar = c = 1$ we have

$$1\text{K} = 8.62 \cdot 10^{-5}\text{eV} , \quad 1\text{eV} = 1.16 \cdot 10^{4}\text{K} \tag{7.5}$$

7.1.3 Conditions for (approximate) thermodynamic equilibrium

One plausible criterion for thermalization is that interactions occur sufficiently frequent. In the expanding FRW universe, there are two relevant time scales. The interaction rate Γ is

$$\Gamma = n\overline{v}\sigma \tag{7.6}$$

where n is the particle density, $\overline{v}$ the average particle velocity (so that $n\overline{v}$ is the particle flux) and σ is the interaction cross section.
This is to be compared to the expansion rate of the universe, the Hubble function

$$H = \frac{\dot{R}}{R} \tag{7.7}$$

As a rule of thumb one expects equilibrium if

$$\Gamma \geq H \tag{7.8}$$

and no equilibrium if

$$\Gamma < H \tag{7.9}$$

A typical feature of cosmology is the existence of threshold temperature, where the interaction rate drops below the value needed to maintain equilibrium. Below the threshold the particle, which have been kept in equilibrium before, decouple. The most important example is 'recombination.'

If the temperature of the universe is to high for the formation of atoms, there are free electrically charged particles which interact with photons and, hence, among themselves. At the end of this period the universe essentially consists off γ, e, p (and n). The reactions

$$\gamma + e \rightarrow \gamma + e \tag{7.10}$$

etc. keep these particles in thermodynamic equilibrium as long as free electrons are available. At the so-called recombination temperature $T_{\rm rec} = 0.3eV = 3375K$, protons and electrons combine into hydrogen atoms, because the average energy of photons becomes to low for the ionization of hydrogen:

$$p + e \to H + \gamma \tag{7.11}$$

From this point on photons and 'baryonic' matter (p, n, e) are decoupled. As we will see the distribution of the photons remains thermal, and its temperature is red shifted to its present value of 2.7 K. Note that the distribution remains thermal despite that the photons essentially do not interact with one another. This is an important exception to the above rule of thumb. We will return to this later in the chapter. The baryonic matter, which is no longer kept in equilibrium by radiation starts to form structures. The reason is that the remaining force between them is gravity, which is purely attractive. Given some initial inhomogenities, these will grow and at some point bound states (the first stars) and other structures form. The standard explanation for the initial inhomogenities is provided by the theory of inflation. These inhomogenities are already present when baryonic matter is still coupled to the photons. At this stage they are very small. They lead to small fluctuations in the photon distribution, which should be observable in the CMB. The actual observation of fluctuations in the CMB, which have precisely the form predicted by inflation, give significant support to the inflationary scenario.

7.1.4 Short thermal history of the universe

We now briefly list the most important events (thresholds) in the history of the universe.

1 The electroweak phase transition at $T \simeq 300$GeV.

   Above this temperature the electroweak theory is in its symmetric phase: the gauge bosons $W^{\pm}, Z^0$, the Higgs doublet and all fermions are strictly massless. (And: quarks and gluons are free particles at these temperatures, see below). We have a hot plasma of massless gauge bosons, fermions and scalars.
   Below this temperature the electroweak theory is in its 'broken' phase: $W^{\pm}, Z^0$, the fermions and the Higgs particle have become massive.
   One might ask why this should be relevant. All masses are below 100 GeV and at temperature much larger than that all particles are effectively massless anyway. The relevance lies in the fact that there is a phase transition. If the phase transition is of first order, then the unbroken and broken phase can coexist as domains separated by domain walls. Such domain walls might lead to observable remnants. (In fact there might be problems if such remnants are predicted but not observed. In particular domain walls potentially lead to stronger inhomogenities than we observe.) Moreover such domain walls might lead to interactions which

violate baryon number conservation. Thus this phase transition might be relevant to understand the matter surplus in the universe. Today it is believed that the phase transition is of first order, but very weak (small discontinuities in thermodynamic quantities).

1a In grand unified theories, there should be a similar phase transition at a temperature of $T \simeq 10^{15}$GeV. At this temperature the GUT group breaks down to the standard model group (meaning that the extra gauge bosons become massive). Again, this transition might lead to observable relicts, for example magnetic monopoles. In fact, the non-observation of magnetic monopoles was histrorically the motivation for the inflationary scenario. (The standard model does not admit magnetic monopoles, while GUT groups do. Therefore there are no magnetic monopoles are produced in the electroweak phase transitions.)

Since GUT groups have interactions which violate baryon number conservation (in particular, they predict proton decay), they might be relevant for explaining baryogenesis (the matter surplus). But so far there is no satisfactory explanation for the observed matter surplus. The problem is that while one can qualitatively obtain matter surplus, it is not large enough (either because baryon number violating processes are to weak or because asymmetries get washed out at later stages. This is, by the way, a problem with non-equilibrium aspects.)

2 The confinement phase transition at $T \simeq 150$MeV.
Above this temperature, one has a quark gluon plasma, while below quarks are confined into hadrons. It is not clear whether this is really a phase transition in the strict sense, it might be a smooth cross over (like vapour/water above the critical point).
For our purposes the only hadrons which are stable at the relevant time scales are protons and neutrons.

3 Nucleosynthesis 10MeV $> T >$ 0.1MeV
Neutrons have a life time of a few minutes and can only survive when bound into nuclei. Primordial nucleosynthesis happens in the interval of temperature indicated above. At the end one gets 75 per cent hydrogen and 24 per cent helium-4.

4 Neutrino and photon decoupling
Above 1 MeV the particle content of the universe is p, n, e, γ, ν which are kept in equilibrium by electromagnetic and weak interactions. (p, n bind to nuclei by strong interactions, see above.) At $T \simeq 1$MeV weak interactions become to weak to keep the neutrinos in equilibrium. This leads to the formation of a (actually three, one for each species) of thermal neutrino background with today should have a temperature of 1.96 K.
Photons decouple when atom form at about $T \simeq 0.31eV$. This leads to the formation of the CMB which is observed today with $T = 2.7K$.

The difference in neutrino and photon temperature is due to the fact that between the two decoupling temperature there is the threshold for the pair production of electrons and positrons,

$$e^- + e^+ \leftrightarrow \gamma + \gamma \tag{7.12}$$

Once atoms have formed and photon have decoupled structure formation starts.
In our discussion we mostly neglected further exotic matter (dark matter) and other relicts. We also neglected the gravitational background

5 Transition from radiation dominance to matter dominance.
Shortly before recombination, the contributions of radiation and matter to the energy density become equal.

7.2 Thermodynamics

Basic equations:
FRW metric

$$ds^2 = -dt^2 + R(t)^2 \, h_{ij}^{(k)} dx^i dx^j \tag{7.13}$$

where $h_{ij}^{(k)}$ is the metric on a maximally symmetric three-dimensional Riemannian space, $k = 0, 1, -1$.

Rewrite energy conservation in a FRW space-time

$$\dot{\rho} + 3\frac{\dot{R}}{R}(\rho + p) = 0 \tag{7.14}$$

as

$$\frac{d}{dt}(\rho R^3) = -p\frac{d}{dt}R^3 \tag{7.15}$$

Take V_0 to be the fixed parametric volume of a co-expanding region G (coexpanding region means that the region has fixed FRW coordinates):

$$V_0 = \int_G d^3x\sqrt{{}^3h} \tag{7.16}$$

The time-dependent, physical volume:

$$V(t) = R(t)^3 \, V_0 \tag{7.17}$$

The energy contained in the region:

$$U = \rho(t)V(t) = \rho(t)R(t)^3 \, V_0 \tag{7.18}$$

It is convenient to consider a unit volume, $V_0 = 1$. Then

$$U = \rho(t)R(t)^3 \text{ and } dU = d(\rho R^3) \tag{7.19}$$

Then (7.15) becomes

$$dU = -pdV \tag{7.20}$$

Comparing to the first theorem of thermodynamics (grand canonical system)

$$dU = -pdV + TdS + \Phi dN \tag{7.21}$$

we see that at least formally the expansion of the universe is adiabatic. This can only be an approximation, but we will see evidence that it is a good approximation most of the time.

The chemical potential

For later use let us review the notion of a chemical potential Φ. In grand canonical systems, where the particle number is allowed to fluctuate, the chemical potential is the intensive quantity conjugate to the (extensive) particle number N. The chemical potential measures particle number fluctuation. In the thermodynamic limit $N \to \infty$. it goes to zero, and grand canonical and canonical ensemble coincide. Chemical potentials are often negligible at high temperatures, too, but we will need them at some places. As a typical example, consider particle interactions of the type

$$A + B \leftrightarrow C + D \tag{7.22}$$

where A, B, C, D are particle species which may or may not be different. Chemical equilibrium is characterized in terms of the Helmholtz free energy,

$$F = U - TS \tag{7.23}$$

$$dF = -SdT - pdV + \Phi dN \tag{7.24}$$

by

$$\frac{\partial F}{\partial N}_{|T,V} = \Phi = 0 \tag{7.25}$$

In our example Φ takes the form

$$\Phi = \Phi_A + \Phi_B - \Phi_C - \Phi_D = 0 \tag{7.26}$$

where Φ_A is the chemical potential of particle species A.

The entropy

For later use, we derive a a formula for the entropy.

Consider first the case of vanishing chemical potential, $\Phi = 0$, corresponding to strict particle number conservation (canonical system). Start from the first theorem

$$dU = TdS - pdV \tag{7.27}$$

and solve for dS:

$$dS = \frac{dU}{T} + \frac{pdV}{T} \tag{7.28}$$

Let us condider T and V as the independent variables, $U = U(T, V)$. Then

$$dS = \frac{1}{T}\frac{\partial U}{\partial T}dT + \frac{1}{T}\left(\frac{\partial U}{\partial V} + p\right)dV \tag{7.29}$$

Evaluate integrability condition:

$$\begin{aligned} ddS = 0 \quad &\Rightarrow \quad \frac{\partial}{\partial V}\left(\frac{1}{T}\frac{\partial U}{\partial T}\right) = \frac{\partial}{\partial T}\left[\frac{1}{T}\left(\frac{\partial U}{\partial V} + p\right)\right] \\ &\Rightarrow \quad T\frac{\partial p}{\partial T}_{|V} = \frac{\partial U}{\partial V}_{|T} + p \end{aligned} \tag{7.30}$$

For the expanding universe we had $U = \rho(t)V = \rho(t)R(t)^3$. This suggests that $\rho = \rho(T)$ is independent of the volume: $\frac{\partial\rho}{\partial V} = 0$. (The whole volume dependence of U resides in the scale factor). For equations of state $p = w\rho$ this implies $p = p(T)$. Using $\frac{\partial U}{\partial V} = \rho$ and $\frac{\partial p}{\partial V} = 0$ we have

$$T\frac{\partial p}{\partial T}_{|V} = \rho + p \tag{7.31}$$

or

$$dp = \frac{\rho + p}{T}dT \tag{7.32}$$

Plugging this back into dS we obtain

$$\begin{aligned} dS &= \frac{d(\rho V)}{T} + \frac{d(pV)}{T} - \frac{V dp}{T} \\ &= \frac{d((\rho + p)V)}{T} - \frac{(\rho + p)V dT}{T^2} \\ &= d\left(\frac{(\rho + p)V}{T}\right) \end{aligned} \tag{7.33}$$

Up to a constant (which we ignore because it spoils extensivity) we find

$$S = \frac{(\rho + p)V}{T} \tag{7.34}$$

and the associated entropy density is

$$s = \frac{S}{V} = \frac{\rho + p}{T} \tag{7.35}$$

To show that S is constant, recall that we found for an expanding FRW universe

$$dU = -pdV \tag{7.36}$$

Rewrite

$$dU = d(\rho V) = -pdV \Rightarrow d((\rho + p)V) = Vdp \tag{7.37}$$

and substitute into

$$\begin{aligned} dS &= d\left(\frac{(\rho+p)V}{T}\right) \\ &= \frac{Vdp}{T} - \frac{(\rho+p)VdT}{T^2} \\ &= \frac{V(\rho+p)dT}{T^2} - \frac{(\rho+p)VdT}{T^2} \\ &= 0 \end{aligned} \tag{7.38}$$

where we used the relation between dp and dT.

For a gas of massless particles we have $p = \frac{1}{3}\rho$ and $\rho \simeq R^{-4}$, while $V = R^3$. Constant entropy requires $T \simeq R^{-1}$. Using the equation of state and the Stefan-Boltzmann law (see below) $\rho = a_0 T^4$, we obtain the entropy density

$$s = \frac{4}{3}\frac{\rho}{T} = \frac{4}{3}a_0 T^3 \tag{7.39}$$

Why does this scaling argument not work for dust, $p = 0$? Dust $\neq$ non-relativistic gas??

7.3 Classical one-particle distributions in curved space-time

Ref: [13]

7.3.1 Distributions

In statistical physics one desribes many-particle systems in terms of phase-space distribtutions, which allow one to compute probability distributions for the macroscopic (thermodynamic) quantities of the system. For thermodynamic equilibrium these can be derived from the second law of thermodynamics (the principle of maximal entropy), or for sufficiently simple systems from the microscopic dynamics, using, for example, the Boltzmann equations. The later type of methods can also be applied to non-equlibrium processes.
One way to obtain the probability distributions for curved space-time is to use the Liouville theorem which expresses the conservation of the phase space

volume, together with information from the Boltzmann equation which we will treat as a 'black box.'
We consider the phase space $\{(x,p)\}$ of relativistic particles in a curved spacetime and focus on the one-particle distributions $f(x,p)$ which tell us the probability of finding a particle with given momentum at a given place. (The modifications due to quantum mechanics will be implemented later). One version of the Liouville theorem is that the one-particle distribution is conserved along trajetories in phase space (curve parameter l):

$$\frac{\partial f}{\partial x^\mu}\frac{dx^\mu}{dl} + \frac{\partial f}{\partial p^\mu}\frac{dp^\mu}{dl} = 0 \tag{7.40}$$

Since $f = \frac{dN}{d\Gamma}$, where dN is the particle number (density) in the phase space volume element $d\Gamma$, this means that if we fix a set of 'particles' dN, then the corresponding phase space volume $d\Gamma$ is conserved. (We certainly expect this behaviour for an equilibrium distribution.)
Consider now particles moving along a congruence of geodesics in a FRW spacetime, with proper time t as curve parameter

$$\frac{\partial f}{\partial t} - R\dot{R}\frac{|\vec{p}|^2}{p^0} - 2\frac{\dot{R}}{R}|\vec{p}|\frac{\partial f}{\partial|\vec{p}|} = 0\ . \tag{7.41}$$

Using the Boltzmann equation it can be shown that the equilibrium distribution for quasistatic, reversible processes takes the form

$$f = e^{\alpha(t)+\beta_\mu(t)p^\mu} \tag{7.42}$$

with arbitrary functions α, β. This is now plugged into the Liouville equation. We use that in a comoving frame $p^\mu = p^0\delta^\mu_0$ and define $\beta = -\beta_0$, so that

$$f = e^{\alpha(t)-\beta(t)p^0} \tag{7.43}$$

We also use the relativistic mass shell condition ($p^0 > 0$ is the energy):

$$\begin{aligned} & p^\mu p_\mu = g_{\mu\nu}p^\mu p^\nu = -(p^0)^2 + R(t)^2|\vec{p}|^2 \\ \Rightarrow\quad & p^0 = \sqrt{m^2 + R^2|\vec{p}|^2} \end{aligned} \tag{7.44}$$

Then the Liouville equation becomes

$$\dot{\alpha} - \dot{\beta}\sqrt{m^2 + R^2|\vec{p}|^2} + \beta R\dot{R}\frac{|\vec{p}|^2}{\sqrt{m^2+|\vec{p}|^2}} = 0 \tag{7.45}$$

The generic solution is $\dot{\alpha} = \dot{\beta} = \dot{R} = 0$, but we want a solution for an expanding universe $\dot{R} \neq 0$. One can find one exact solution (radiation) and one approximate solution (non-relativistic matter).

1. Radiation: $m = 0$. Then the equation

$$\dot{\alpha} + |\vec{p}|(\dot{\beta}R - \beta\dot{R}) = 0 \tag{7.46}$$

is solved by

$$\alpha = \alpha_0 \ , \quad \beta = \beta_0 R(t) \tag{7.47}$$

where α_0, β_0 are constant.

2. Non-relativistic matter, $m \to \infty$.
   Expand in

$$\frac{R^2|\vec{p}|^2}{m^2} \ll 1 \tag{7.48}$$

   Expansion of the square root gives

$$\begin{aligned} & \dot{\alpha} - \dot{\beta} m \left(1 + \frac{1}{2}\frac{R^2|\vec{p}|^2}{m^2} + \cdots\right) + \beta R\dot{R}\left(1 - \frac{1}{2}\frac{R^2|\vec{p}|^2}{m^2} + \cdots\right) = 0 \\ \Rightarrow \quad & \dot{\alpha} - m\dot{\beta} + \frac{|\vec{p}|^2}{m^2}\left(-\frac{1}{2}\dot{\beta}R^2 + \beta\dot{R}R\right) + \mathcal{O}[(\tfrac{R|\vec{p}|}{m})^3] = 0 \end{aligned} \tag{7.49}$$

   This is solved by

$$\alpha = m\beta + \alpha_0 \ , \quad \beta = \beta_0 R^2 \tag{7.50}$$

Now we turn to the physical interpretation.

7.3.2 Thermodynamics of the radiation dominated universe

Start from the equilibrium distribution

$$f = e^{\alpha_0 - \beta(t)p^0} = e^{\alpha_0 - \beta_0 R(t) p^0} \tag{7.51}$$

Note that $\beta \sim R$, more precisely

$$\frac{\beta(t)}{\beta(t_0)} = \frac{R(t)}{R(t_0)} \tag{7.52}$$

where $R(t_0) = 1$ relates the normalizations of $\beta(t)$, $R(t)$. Since p^0 is the energy E, we should identify

$$e^{-\beta(t)p^0} = e^{-\frac{E}{k_B T}} \Rightarrow \beta(t) = \frac{1}{k_B T(t)} \tag{7.53}$$

where k_B is the Boltzmann constant and $T = T(t)$ is the temperature of the universe at the time t. We will often set $k_B = 1$ in the following.
Temperature scales with the inverse scale factor, $T \sim R^{-1}$, or more precisely:

$$\frac{T(t)}{T(t_0)} = \frac{\beta(t_0)}{\beta(t)} = \frac{R(t_0)}{R(t)} \tag{7.54}$$

Observe that $T \sim R^{-1}$ is consistent with the Stefan-Boltzmann law for the energy density of black body radiation,

$$\rho = a_0 T^4 \tag{7.55}$$

since for radiation

$$\rho \sim R^{-4} \sim T^4 \tag{7.56}$$

We also note that if $R \to 0$ for $t \to 0$ then $T \to \infty$. Thus the big bang is 'hot'.

Relation to the redshift:

$$1 + z = \frac{R(t)}{R(t_0)} = \frac{T(t_0)}{T(t)} \tag{7.57}$$

7.3.3 Thermodynamics of a matter dominated universe

Equilibrium distribution:

$$f = e^{\alpha(t)-\beta(t)p^0} = e^{(m\beta_0 R^2+\alpha_0)-\beta_0 R^2 p^0} \tag{7.58}$$

matching again $\beta(t) = \frac{1}{T}$ we have $T \sim R^{-2}$ or

$$\frac{R(t)^2}{R(t_0)^2} = \frac{T(t_0)}{T(t)} = (1+z)^2 \tag{7.59}$$

Observe that in an expanding universe matter ($T \sim R^{-2}$) cools down faster than radiation ($T \sim R^{-1}$).

Let us check that $T \sim R^{-2}$ is consistent with $\rho \sim R^{-3}$, which requires $\rho \sim T^{\frac{3}{2}}$. The energy density of non-relativistic particles is dominated by their mass, kinetic energy can be neglected:

$$\rho = mn \tag{7.60}$$

where n is particle density. The particle density is given by the Boltzmann distribution (see below)

$$n = g\left(\frac{mT}{2\pi}\right)^{\frac{3}{2}} e^{-\frac{m-\Phi}{T}} \tag{7.61}$$

where g counts the number of internal states of the particle (spin), and we assumed $T \ll m - \Phi$ (non-relatistic limit).
To lowest order we have

$$\rho \sim n \sim T^{\frac{3}{2}} \tag{7.62}$$

as needed. (We used that $\exp(-\frac{1}{T})$ varies slowly in T for T small, see below.)

7.4 Quantum mechanical one-particle distributions

In the previous section we used classical statistical physics. In quantum theory particles do not obey Maxwell-Boltzmann statistics, but Bose-Einstein statistics (bosons, integer spin) or Fermi-Dirac statistics (fermions, half-integer spin).

In the following we do not derive them for curved space-time, but use that the physics of a FRW universe in comoving coordinates looks like flat space physics. Therefore we can use the standard one-particle distributions of equilibrium quantum statistical physics, except that we have to take into account that the physical volumes scales with time as $V \simeq R(t)^3$.
In kinetic equilibrium (fluctuations around an equilibrium temperature) we have the standard Bose-Einstein distribution $\varepsilon = 1$ for bosons and Fermi-Dirac distribution $\varepsilon = -1$ for fermions:

$$f(\vec{p}) = \frac{1}{\exp\left(\frac{(E-\Phi)}{T}\right) - \varepsilon} \tag{7.63}$$

where E is the energy and Φ is the chemical potential. (In chemical equilibrium one has the additional condition that the total chemical potential of the reaction under consideration vanishes, corresponding to fluctuations around an average particle number. The case with sharp particle numbers is obtained by setting Φ identically zero.)
The (expectation values of) thermodynamical quantities can then be computed:
Particle density:

$$n = \frac{g}{(2\pi)^3} \int d^3\vec{p}\, f(\vec{p}) \tag{7.64}$$

Energy density:

$$\rho = \frac{g}{(2\pi)^3} \int d^3\vec{p}\, E(\vec{p})\, f(\vec{p}) \tag{7.65}$$

Pressure

$$p = \frac{g}{(2\pi)^3} \int d^3\vec{p}\, \frac{|\vec{p}|^2}{3E}\, f(\vec{p}) \tag{7.66}$$

where g = number of internal degrees of freedom (spin).

$$E = p^0 = \sqrt{m^2 + |\vec{p}|^2} \tag{7.67}$$

is the energy. To be precise, this is the flat-space form of the dispersion relation for a relativistic particle. As we saw above, in curved space-time we have $|\vec{p}|^2 \to R^2(t)|\vec{p}|^2$, so that $E = E(t)$ becomes a function of time. For simplicity, we will set $R(t) = 1$ in the following, which we can do when considering time scales which are short compared to the expansion rate (a momentary quasistatic equilibrium state). The consequences of the time-dependence of E will be studied in a second step.
The above distributions are independent of x^μ, as a consequence of the translation invariances of Minkowski space (which we take as the local approximation of our FRW space-time). Rotation invariance implies that the momentum dependence is only through $|\vec{p}|$, which is true for the explicit distributions specified above. The 'spherical' integration in momentum space can be performed exactly.
We use

$$d^3\vec{p} = |\vec{p}|^2 d|\vec{p}| d^2\Omega \tag{7.68}$$

where $d^2\Omega$ is the volume element of unit-two sphere,

$$\int_{S^2} d^2\Omega = 4\pi \tag{7.69}$$

Then

$$\begin{aligned} n &= \frac{g}{2\pi^2}\int d|\vec{p}|\frac{|\vec{p}|^2}{\exp\left(\frac{(E-\Phi)}{T}\right)-\varepsilon} && (7.70)\\ \rho &= \frac{g}{2\pi^2}\int d|\vec{p}|\frac{|\vec{p}|^2 E}{\exp\left(\frac{(E-\Phi)}{T}\right)-\varepsilon} && (7.71)\\ p &= \frac{g}{6\pi^2}\int d|\vec{p}|\frac{|\vec{p}|^2|\vec{p}|^2}{E\exp\left(\frac{(E-\Phi)}{T}\right)-\varepsilon} && (7.72)\\ & && (7.73) \end{aligned}$$

Using that $E = p^0 = \sqrt{m^2 + |\vec{p}|^2}$ we can rewrite the radial integration in terms of the energy,

$$|\vec{p}|^2 d|\vec{p}| = \sqrt{E^2 - m^2}EdE \tag{7.74}$$

$$\begin{aligned} n &= \frac{g}{2\pi^2}\int_m^\infty \frac{\sqrt{E^2-m^2}}{\exp\left(\frac{(E-\Phi)}{T}\right)-\varepsilon}EdE && (7.75)\\ \rho &= \frac{g}{2\pi^2}\int_m^\infty \frac{\sqrt{E^2-m^2}}{\exp\left(\frac{(E-\Phi)}{T}\right)-\varepsilon}E^2dE && (7.76)\\ p &= \frac{g}{6\pi^2}\int_m^\infty \frac{(E^2-m^2)^{\frac{3}{2}}}{\exp\left(\frac{(E-\Phi)}{T}\right)-\varepsilon}E^2dE && (7.77) \end{aligned}$$

An analytic evaluation of the remaining integral is possible in certain interesting asymptotic regimes.

The ultra-relativistic limit

Consider temperatures, which are large compared to both the mass $T \gg m$ and the chemical potential $T \gg \Phi$. Then

$$\begin{aligned} n &= \frac{g}{2\pi^2}\int d|\vec{p}|\frac{|\vec{p}|^2}{\exp\left(\frac{E}{T}\right)-\varepsilon} && (7.78)\\ \rho &= \frac{g}{2\pi^2}\int d|\vec{p}|\frac{|\vec{p}|^2 E}{\exp\left(\frac{E}{T}\right)-\varepsilon} && (7.79)\\ p &= \frac{g}{6\pi^2}\int d|\vec{p}|\frac{|\vec{p}|^2|\vec{p}|^2}{E\exp\left(\frac{E}{T}\right)-\varepsilon} && (7.80)\\ & && (7.81) \end{aligned}$$

where $E = |\vec{p}|$. Setting

$$\frac{E}{T} = \frac{|\vec{p}|}{T} =: y \tag{7.82}$$

we find

$$n = \frac{g}{2\pi^2} T^3 \int_0^\infty dy \frac{y^2}{e^y - \varepsilon} \tag{7.83}$$

$$\rho = \frac{g}{2\pi^2} T^4 \int_0^\infty dy \frac{y^3}{e^y - \varepsilon} \tag{7.84}$$

$$p = \frac{g}{6\pi^2} T^4 \int_0^\infty dy \frac{y^3}{e^y - \varepsilon} \tag{7.85}$$

$$\tag{7.86}$$

Observe that we have derived the equation of state of a gas of massless particles

$$p = \tfrac{1}{3}\rho \tag{7.87}$$

The integrals can be evaluated:

$$\int_0^\infty dy \frac{y^m}{e^y - \varepsilon} = \begin{cases} \zeta(m+1)\Gamma(m+1) & \text{for } \varepsilon = +1 \\ \frac{2^m - 1}{2^m}\zeta(m+1)\Gamma(m+1) & \text{for } \varepsilon = -1 \end{cases} \tag{7.88}$$

See appendix A.3, for the details.

Thus

$$n = \begin{cases} \frac{g\zeta(3)}{\pi^2} T^3 & \text{for } \varepsilon = +1 \\ \frac{3}{4}\frac{g\zeta(3)}{\pi^2} T^3 & \text{for } \varepsilon = -1 \end{cases} \tag{7.89}$$

$$\rho = \begin{cases} \frac{g\pi^2}{30} T^4 & \text{for } \varepsilon = +1 \\ \frac{7}{8}\frac{g\pi^2}{30} T^4 & \text{for } \varepsilon = -1 \end{cases} \tag{7.90}$$

where we used $\zeta(4) = \frac{\pi^4}{90}$. $\zeta(3)$ has the approximate value $\zeta(3) = 1.20206\ldots$.

For photons ($g = 2$) we recognize the Stefan Bolzmann law:

$$\rho = \frac{\pi^2}{15} T^4 = a_0 T^4 \tag{7.91}$$

Reconstructing $\hbar, c, k_B$:

$$a_0 = \frac{\pi^2}{15} \frac{k_B^4}{\hbar^3 c^3} \tag{7.92}$$

We note the following useful relations between bosonic and fermionic ultrarelativistic denstities:

$$\frac{n_B}{n_F} = \frac{3}{4}\,, \quad \frac{\rho_B}{\rho_F} = \frac{7}{8} = \frac{p_B}{p_F} = \frac{s_B}{s_F} \tag{7.93}$$

(where we used that the entropy density is $s = \frac{\rho + p}{T}$).

The non-relativistic limit

In the non-relativistic limit, we have $m \gg T$ and $m \gg p$ so that $E = m + \frac{1}{2m}|\vec{p}|^2$. Thus

$$\begin{aligned} n &= \frac{g}{2\pi^2}\int d|\vec{p}|\frac{|\vec{p}|^2}{e^{\frac{E-\Phi}{T}} - \varepsilon} && (7.94)\\ &\approx \frac{g}{2\pi^2}\int d|\vec{p}||\vec{p}|^2 e^{-\frac{(m-\Phi)+\frac{|\vec{p}|^2}{2m}}{T}} && (7.95)\\ &= \frac{ge^{-\frac{m-\Phi}{T}}}{2\pi^2}(2mT)^{\frac{3}{2}}\int_0^\infty dy y^2 e^{-y^2} && (7.96)\\ &= ge^{-\frac{m-\Phi}{T}}\left(\frac{mT}{2\pi}\right)^{\frac{3}{2}} && (7.97)\end{aligned}$$

Here we set $|\vec{p}| = \sqrt{2mT}y$ and used

$$\int_0^\infty dy y^2 e^{-y^2} = -\frac{d}{d\alpha}_{|\alpha=1}\int_0^\infty dy e^{-\alpha y^2} = -\frac{d}{d\alpha}_{|\alpha=1}\frac{\sqrt{\pi}}{2\alpha} = \frac{\sqrt{\pi}}{4} \qquad (7.98)$$

Note that for $m \gg T$ the exponential is a slowly varying function. If we send $\frac{m-\Phi}{T} \to \infty$, while keeping m, Φ fixed we have

$$\frac{d}{dT}e^{-\frac{m-\Phi}{T}} = \frac{m-\Phi}{T^2}e^{-\frac{m-\Phi}{T}} = \frac{1}{m-\Phi}\left(\frac{m-\Phi}{T}\right)^2 e^{-\frac{m-\Phi}{T}} \to 0 \qquad (7.99)$$

Since

$$\lim_{x\to\infty} x^m e^{-x} = 0 \qquad (7.100)$$

Since the exponential varies slowly for small temperatures (compared to the mass), the essential temperature dependence is $n \sim T^{\frac{3}{2}}$. (This was used above to show that for a matter dominated universe $\rho \simeq R^{-3}$ and $T \simeq R^{-2}$ are compatible.)

Important observation:
If one has relativistic and non-relativistic particle species in equilibrium, then the distributions of the non-relativistic particles are exponentially suppressed. Therefore it is a good approximation to only consider the relativistic particles.

Several relativistic (and non-relativistic) species in equilibrium

As a slight generalization, consider several particle species in equilibrium, where we allow that each species has a different temperature T_i:

$$\rho = \sum \rho_i = \sum_i \frac{g_i}{2\pi}\int_{m_i}^\infty \frac{\sqrt{E^2 - m_i^2}}{\exp\left(\frac{(E-\Phi_i)}{T_i}\right) - \varepsilon_i}E^2 dE \qquad (7.101)$$

Introduce variables

$$x_i = \frac{m_i}{T_i}\ , \quad y_i = \frac{\Phi_i}{T_i}\ , \quad u_i = \frac{E}{T_i} \tag{7.102}$$

Since u_i is an integration variable we can rename it u for each integral. It is convenient to express the temperatures T_i in terms of their relation to the photon temperature $T = T_\gamma$.

$$\begin{aligned}\rho = \sum_i \rho_i &= \sum_i (T_i)^4 \frac{g_i}{2\pi^2} \int_{x_i}^{\infty} \frac{\sqrt{u - x_i}}{\exp(u - y_i) - \varepsilon_i} u^2 du \\ &= T^4 \sum_i \frac{T_i^4}{T^4} \frac{g_i}{2\pi^2} \int_{x_i}^{\infty} \frac{\sqrt{u - x_i}}{\exp(u - y_i) - \varepsilon_i} u^2 du\end{aligned}$$

Since non-relativistic particles are exponentially suppressed, we obtain a simplified expression by only summing over the ultra-relativistic particles, $m_i \ll T$, and using the asymptotic formulae for the integrals. Then we obtain

$$\rho = \frac{\pi^2}{30} T^4 \left(\sum_{i \in RB} g_i \frac{T_i^4}{T^4} + \frac{7}{8} \sum_{i \in RF} g_i \frac{T_i^4}{T^4} \right) \tag{7.103}$$

Note the relative factor between relativistic bosons (RB) and relativistic fermions (RF). Defining a temperature-dependent effective number of degrees of freedom

$$g_* = \sum_{i \in RB} g_i \frac{T_i^4}{T^4} + \frac{7}{8} \sum_{i \in RF} g_i \frac{T_i^4}{T^4} \tag{7.104}$$

we have formally a one-particle formula

$$\rho = \frac{\pi^2}{30} T^4 g_* \tag{7.105}$$

Similar expression can be found for the particle density n and the pressure p. In the relativistic limit we have $p = \frac{1}{3}\rho$. In this limit the entropy density is

$$s = \sum_i s_i = \sum_i \frac{\rho_i + p_i}{T_i} = \frac{2\pi^2}{45} T^3 g_{*,S} \tag{7.106}$$

where

$$g_{*,S} = \sum_{i \in RB} g_i \left(\frac{T_i}{T} \right)^3 + \frac{7}{8} \sum_{i \in RF} g_i \left(\frac{T_i}{T} \right)^3 \tag{7.107}$$

Observe that $g_* = g_{*,S}$ if $T_i = T$ for all i, but in general they are different.

7.5 Thresholds and decoupling

7.5.1 Thresholds

The effective number of degrees of freedom is temperature dependent even if the temperature of all species are equal, $T_i = T$. The reason is that it jumps whenever a species becomes non-relativistic. Consider a particle X with mass m_X. There are three regimes:

1. $T = T_1 \gg m_X$.
   X is ultrarelativistic. The total entropy is
   $$S_1 = \frac{2\pi^2}{45} T_1^3 g_{*,S,1} R_1^3 \tag{7.108}$$
   where X is included in $g_{*,S,1}$.

2. $T \approx m_X$
   At this temperature X becomes non-relativistic. We don't have an easy way to describe this explicitly so we proceede immediately to the next regime.

3. $T \ll m_X$.
   X is non-relativistic and does not contribute (significantly) to the entropy
   $$S_2 = \frac{2\pi^2}{45} T_2^3 g_{*,S,2} R_2^3 \tag{7.109}$$
   where $g_{*,S,2}$ only gets contributions from the remaining relativistic species. Since the expansion is adiabatic, entropy is conserved, $S_1 = S_2$ and therefore
   $$\frac{T_1}{T_2} = \frac{R_2}{R_1} \left(\frac{g_{*,S,2}}{g_{*,S,1}} \right)^{1/3} \tag{7.110}$$

Therefore thresholds where particles become massless modify the slope in the relation $T \sim R^{-1}$. The entropy of the non-relativistic species is transferred to the remaining relativistic species. As a consequence the decrease in temperature slows down.

7.5.2 Equilibrium from gauge interactions

If a reaction is necessary to maintain equilibrium, then the reaction rate Γ must be at least of order of the expansion rate H:

$$\begin{aligned} \Gamma \geq H \quad &\Rightarrow \quad \text{equilibrium} \\ \Gamma < H \quad &\Rightarrow \quad \text{no equilibrium} \end{aligned} \tag{7.111}$$

Remark 1: This is a rule of thumb, the strict analysis is based on the Boltzmann equation

Remark 2: There are examples where a distribution remains thermal even in absence of interactions. The standard example is the decoupling of massless particles.
Remark 3: Related to the average free travelling length λ: $\Gamma \propto \lambda^{-1}$, thus $\lambda \sim H^{-1}$ is the threshold. If λ becomes larger then the characteristic length scale of expansion, equilibrium cannot be maintained.

The expansion rate

$$H = \frac{\dot{R}}{R} \tag{7.112}$$

can be rewritten in terms of the temperature for a radiation dominated universe, as follows:

$$\frac{T}{T_1} = \frac{R_1}{R} \Rightarrow \frac{\dot{T}}{T_1} = -\frac{R_1}{R^2}\dot{R} = -\frac{R_1}{R}H \tag{7.113}$$

where $T = T(t)$, $T_1 = T(t_1)$, etc. Setting $t_1 = t$:

$$\frac{\dot{T}}{T} = -\frac{\dot{R}}{R} = -H \tag{7.114}$$

The reaction rate is proportional to the particle density n and the cross section σ. (The temperature dependence of the cross section may be obtained by the leading behaviour in energy in the regime under consideration.)

$$\Gamma \simeq n\sigma \tag{7.115}$$

(We consider relativistic particles, $\overline{v} = c = 1$.
Consider gauge interactions mediated by massless spin 1 particles. The dominant processes which keep equilibrium are scattering processes between two charged particles

$$A + B \to C + D \tag{7.116}$$

From the leading energy behaviour of Coulomb-like scattering:

$$\sigma \simeq \alpha^2 T^{-2} \tag{7.117}$$

Here α is the 'finestructure constant' of the interaction under consideration:

$$4\pi\alpha = g^2 \tag{7.118}$$

where g is the coupling constant in the Lagrangian. Each vertex contributes a factor g, there are two in each diagram, and the cross section is proportional to the transition amplitude squared.

Examples:
Electromagnetic interactions:

$$\begin{aligned} e^- + e^- &\leftrightarrow e^- + e^- \\ e^+ + e^- &\leftrightarrow e^+ + e^- \\ e^+ + e^- &\leftrightarrow \gamma + \gamma \\ e^- + p^+ &\leftrightarrow e^- + p^+ \end{aligned} \tag{7.119}$$

Weak interactions:

$$e^- + \nu_e \quad \leftrightarrow \quad e^- + \nu_e \tag{7.120}$$

Strong interactions:

$$q + q' \quad \leftrightarrow \quad q + q' \tag{7.121}$$

For massless particles the particle density behaves like $n \sim T^3$, because:

$$n = \frac{\rho}{\hbar\overline{\omega}} \sim \frac{a_0 T^4}{T} \sim T^3 \tag{7.122}$$

Here we used that for massless particles we have the Stefan Boltzmann law $\rho \sim T^4$, and that the average energy $\hbar\overline{\omega}$ of particles is the temperature $\hbar\overline{\omega} \sim T$. The reaction rate for scattering mediated by massless gauge bosons:

$$\Gamma \sim \alpha^2 T \tag{7.123}$$

To estimate the expansion rate, we consider a radiation dominated universe (without cosmological constant) and flat spatial sections. Then:

$$R(t) = \sqrt{2} b^{\frac{1}{4}} \sqrt{t} \tag{7.124}$$

where

$$b = \frac{8\pi G_N}{3} \rho_1 R_1^4 = \text{const.} \tag{7.125}$$

Then

$$H = \frac{\dot{R}}{R} \sim \frac{1}{t} \sim \frac{\sqrt{b}}{R^2} \sim \frac{\sqrt{G_N}\sqrt{\rho} R^2}{R^2} \sim \sqrt{G_N} T^2 \sim \frac{T^2}{m_{\rm Pl}} \tag{7.126}$$

where we used $\rho \sim T^4$ and $\sqrt{G_N} \sim m_{\rm Pl}^{-1}$.
The condition for equilibrium is

$$\frac{\Gamma}{H} \sim \frac{\alpha^2 T}{\frac{T^2}{m_{\rm Pl}}} = \frac{\alpha^2 m_{\rm Pl}}{T} \geq 1 \Rightarrow T \leq \alpha^2 m_{\rm Pl} \approx \alpha^2 10^{19}\text{GeV} \tag{7.127}$$

Example: the universe at $T \geq 300$GeV.

Above the electroweak phase transition at $T \simeq 300$GeV the universe is a hot plasma of gauge bosons, fermions and scalars. The massless gauge bosons mediate strong and electroweak interactions with couplings $\alpha_i = \frac{g_i^2}{4\pi}$, $i = 1, 2, 3$. Due to quantum corrections, the effective couplings are energy dependent (so-called running couplings). For example $\alpha_{\rm el-mg} \approx \frac{1}{137}$ at low energies, but it rises towards higher energies. In contrast, the effective coupling of strong interactions

decreases at high energies. At around 10^{16} GeV, the three couplings are approximately equal, $\alpha_i \approx \frac{1}{25}$. With this coupling one finds as maximal temperature for maintaining equilibrium of $T \approx 10^{16}$GeV. Therefore gauge interactions can bring and keep the universe into equilibrium for $T \leq 10^{16}$GeV, while it is not clear whether the universe was in equilibrium at earlier times/higher temperature.
In the range $10^{16}\text{GeV} > T > 300\text{GeV}$ we have equilibrium and the universe is a hot plasma of gauge bosons $W^{\pm}, Z^0, \gamma$, 8 gluons, two complex Higgs scalars (unbroken phase) and 3 generations (and antigenerations) of quarks (3 colours) and leptons. The effective number of degrees of freedom is

$$g_* = 2_{\text{Spin}} \cdot (3+1+8) + 4 + \frac{7}{8} \cdot 2_{\text{spin}} \cdot 2_{\text{T/AT}} \cdot 3_{\text{Gen.}} \cdot 2_{\text{weakdoubl/singl.}} \cdot (1+3_{\text{color}}) = 108 \tag{7.128}$$

If there was no right-handed neutrino and left-handed anti-neutrino, we would have to subtract $2 \cdot \frac{7}{8}$, resulting in $g_* = 106, 25$.

Remark: It can be shown that a plasma at very high temperatures behaves like an ideal gas [3]. Thus the details of the interactions do not enter into the equilibrium distributions. The only relevant property of interactions is that they must occur frequently enough to maintain equilibrium. This condition fails at trans-Planckian tempeatures $T > 10^{16}$GeV.

Let us next consider the opposite extreme of gauge interactions mediated by very massive gauge bosons with mass m_X. Here the exchange of gauge bosons with coupling g at each vertex can be replaced by an effective vertex with coupling

$$G_X = \frac{\alpha}{m_X^2} \sim \frac{g^2}{m_X^2} \tag{7.129}$$

The cross section has the following temperature dependence for such interactions:

$$\sigma \sim G_X^2 T^2 \tag{7.130}$$

We consider the case where the charged particles which are kept in equilibrium by the interaction are still light (ultrarelativistic). Then $n \sim T^3$ and

$$\Gamma \sim n\sigma \sim T^3 \left(\frac{\alpha}{m_X^2}\right)^2 T^2 = \left(\frac{\alpha}{m_X^2}\right)^2 T^5 \tag{7.131}$$

(We assume that the universe is radiation dominated.) Condition for equilibrium:

$$\frac{\Gamma}{H} \sim \frac{\frac{\alpha}{m_X^2} T^5}{\frac{T^2}{m_{\text{Pl}}}} = \left(\frac{\alpha}{m_X^2}\right)^2 m_{\text{Pl}} T^3 \geq 1 \tag{7.132}$$

or

$$T \geq \left(\frac{\alpha}{m_X^2}\right)^{-\frac{2}{3}} m_{\text{Pl}}^{-\frac{1}{3}} \tag{7.133}$$

Example: electroweak interactions below $T \simeq 300$ GeV

Below the electroweak phase transition the W and Z bosons aquire a mass of roughly $m_X \approx 100\text{GeV}$. At non-relativistic energies and temperatures $T \ll m_X$, the W and Z exchange can be replaced by the effective Fermi interaction between 4 fermions with coupling G_F. In this case the minimal temperature for maintaining equlibrium through weak interactions is

$$T \geq \left(\frac{m_X^2}{100\text{GeV}}\right)^{\frac{4}{3}} \text{MeV} \approx 1\text{MeV} \tag{7.134}$$

Below this temperature, electrically charged particles like electrons still remain in equilibrium, through their interactions with photons:

$$e^{\pm} + e^{\pm} \quad \leftrightarrow \quad e^{\pm} + e^{\pm} \tag{7.135}$$

However, neutral particles (neutrinos), can only remain in equilibrium through interactions mediated by $W^{\pm}, Z^0$:

$$\begin{aligned} e^{\pm} + \nu_e \quad &\leftrightarrow \quad e^{\pm} + \nu_e \\ e^{+} + e^{-} \quad &\leftrightarrow \quad \nu_e + \overline{\nu}_e \end{aligned} \tag{7.136}$$

Since these interactions 'freeze out' at $T \simeq 1\text{MeV}$, the neutrinos decouple from the other particles at this temperature.
We will now analyse what happens with the statistical distributions of particles which decouple from equilibrium, because the reaction rate drops below expansion rate. Two asymptotic regimes can be treated easily: the decoupling of a species which is ultra-relativistic at decoupling (such as neutrinos) and the decoupling of a non-relativistic species. In both cases the distributions remain thermal (despite the lack of interactions) but is redshifted through the expansion. For neutrinos and photons this leads to the predictions of thermal backgrounds which can be observed today, and which have been observed for photons (CMB).

7.5.3 Decoupling of a ultra-relativistic species

Assume that the decoupling temperature is so high that the species which decouples can be treated as massless, $T_D \gg m$. The distribution of the decoupling particle at the decoupling time is

$$f(\vec{p}, t_D) = \frac{1}{\exp\left(\frac{E}{T_D}\right) - \varepsilon} \tag{7.137}$$

After decoupling, the energy of each massless particle is redshifted by the expansion rate,

$$E(t) = E(t_D)\frac{R(t_D)}{R(t)} \tag{7.138}$$

(c.f. the gravitational redshift for photons), while the particle density decreases as

$$n \sim R^{-3} \tag{7.139}$$

(particle number is conserved, since we assume absence of (relevant) interactions after t_D). Since the distribution is given by

$$f = \frac{\Delta n}{\Delta p^1 \Delta p^2 \Delta p^3} \tag{7.140}$$

where $n \sim R^{-3}$ and $p^i \sim R{-}1$, it is frozen in: $f(\vec{p}, t) = f(\vec{p}_D, t_D)$, for $t > t_D$, where the momenta are related by the redshift factor. Then $f(\vec{p}, t)$ is still a thermal distribution,

$$f(\vec{p}, t) = f(\vec{p}_D, t_D) = f(\vec{p}\frac{R}{R_D}, t_D) = \frac{1}{\exp\left(\frac{ER}{T_D R_D}\right) - \varepsilon} = \frac{1}{\exp\left(\frac{E}{T}\right) - \varepsilon} \tag{7.141}$$

with the temperature red-shifted according to

$$T = T_D \frac{R_D}{R} \tag{7.142}$$

which is precisely the red-shift factor for a radiation dominated universe. Note that the distribution remains thermal, despite that there are essentially no interactions. The reason is that the form of the distribution is preserved by the simple scaling of the momenta caused by the red-shift (in the ultra-relativistic regime).

7.5.4 Decoupling of a non-relativistic species

Consider now the case of a particle decoupling from equilibrium at $T_D \ll m$. For each particle the momentum $\vec{p}$ scales like

$$|\vec{p}|(t) = |\vec{p}|_D \frac{R_D}{R} \tag{7.143}$$

through the gravitational red-shift. Since the particle is non-relativistic, the kinetic energy takes the form

$$E_K(t) = \tfrac{1}{2m}|\vec{p}|^2(t) = E_{k,D}\frac{R_D^2}{R^2} \sim R^{-2} \tag{7.144}$$

and scales with the appropriate power for a matter-dominated universe. Recall the non-relativistic limit of the momentum distribution

$$f(\vec{p}, t) = \exp\left(\frac{-(m - \phi(t)) - \frac{1}{2m}|\vec{p}|^2(t)}{T(t)}\right) \tag{7.145}$$

At decoupling the distribution is

$$f(\vec{p}, t_D) = \exp\left(-\frac{m - \phi_D}{T_D}\right)\exp\left(-\frac{\frac{1}{2m}|\vec{p}|^2}{T_D}\right) \tag{7.146}$$

Again we expect that the distribution is frozen in $f(\vec{p}, t) = f(\vec{p}_D, t_D)$ for $t > t_D$, with momenta related by the redshift. We need to assume that the chemical potential remains constant, and we will see below that this is consistent.
Then we find

$$f(\vec{p}, t) = \exp\left(-\frac{m - \phi_D}{T_D}\right) \exp\left(-\frac{\frac{1}{2m}|\vec{p}|^2 \frac{R_D^2}{R^2}}{T_D}\right) = \exp\left(-\frac{m - \phi_D}{T_D}\right) \exp\left(-\frac{|\vec{p}|^2}{2mT}\right) \tag{7.147}$$

where

$$T = T_D \frac{R_D^2}{R^2} \tag{7.148}$$

This is the scaling relation $T \sim R^{-2}$ and we see that the Boltzmann factor scales such that the distribution has the standard form of a non-relativistic thermal distribution, except for the prefactor. But we can re-interprete our result as a non-relativistic thermal distribution with time-depenendent chemical potential:

$$f(\vec{p}, t) = \exp\left(-\frac{m - \phi(t)}{T}\right) \exp\left(-\frac{|\vec{p}|^2}{2mT}\right) \tag{7.149}$$

where

$$\frac{m - \phi_D}{T_D} = \frac{m - \phi(t)}{T} \Leftrightarrow \phi(t) = m + (\phi_D - m)\frac{T}{T_D} \tag{7.150}$$

This is a very special time-dependence, because the prefactor of the Boltzmann factor is constant (the chemical potential compensates the time-dependence of the temperature).
Since particle number is conserved, the particle density decreases as

$$n \sim R^{-3} \tag{7.151}$$

This result follows from geometry, and we would like to check that it is consistent with thermodynamics. (We also need it to have a constant phase space distribution, which we assumed above).

We can compute n from the momentum distribution:

$$\begin{aligned} n &= \frac{g}{2\pi^2} e^{-\frac{m-\phi}{T}} \int d|\vec{p}||\vec{p}|^2 \exp\left(-\frac{|\vec{p}|^2}{2mT}\right) \\ &= \frac{g}{2\pi^2} e^{-\frac{m-\phi_D}{T_D}} \int d|\vec{p}||\vec{p}|^2 \exp\left(-\frac{|\vec{p}|^2}{2mT}\right) \\ &= g e^{-\frac{m-\Phi_D}{T_D}} \left(\frac{mT}{2\pi}\right)^{\frac{3}{2}} \qquad (7.152) \\ &= g e^{-\frac{m-\phi(t)}{T}} \left(\frac{mT}{2\pi}\right)^{\frac{3}{2}} \qquad (7.153) \end{aligned}$$

Since the prefactor is constant, we have

$$n \sim T^{\frac{3}{2}} \tag{7.154}$$

Using the scaling $T \sim R^{-2}$ we find $n \sim R^{-3}$, and the geometrical and thermodynamical computation of the particle density are consistent.

7.6 Selected applications

We now apply the results of the last section to several processes occuring in cosmology.

7.6.1 The photon background

Charged particle remain in equlibrium up to the point where atoms form and matter becomes effectively neutral. At this point matter means protons, neutrons and electrons. Quarks have now been confined into hadrons, and protons and neutrons are the only stable hadrons at the time scale under consideration. Since electrons are much lighter, they dominate the matter distribution and we focus on them. The dominant process for maintaining equilibrium is

$$\gamma + e \to \gamma + e \tag{7.155}$$

Photon decoupling occurs when the density of free electrons sinks below a certain threshold.
The most relevant process for eliminanating free electrons is (re)combination/ionization:

$$p + e \leftrightarrow H + \gamma \tag{7.156}$$

In the most naive approximation, atoms forms at some temperature and eliminate the free electrons. One might expect that ionization stops when the temperature reaches the ionization energy $E_{\text{ion}} = 13.6\text{eV}$. However, there is a significant residual ionization below this temperature. Ionization only stops (effectively) at the recombination temperature

$$T_{\text{rec}} = 0.31\text{eV} = 3.6 \cdot 10^3 K \tag{7.157}$$

(This will be discussed in more detail below).
Since the photon is massless, its distribution remains thermal, with temperature

$$\frac{T}{T_{\text{rec}}} = \frac{R_{\text{rec}}}{R} = (1 + z_{\text{rec}})^{-1} \tag{7.158}$$

This is the cosmic mircowave background or CMB. The value observed today is

$$T \approx 2.73K \tag{7.159}$$

The redshift at recombination time is roughly

$$z_{\mathrm{rec}} \simeq 1300 \tag{7.160}$$

and the recombination time is roughly

$$t_{\mathrm{rec}} \simeq 10^5 y \tag{7.161}$$

Note that the details are more complicated than we assumed here, because recombination (formation of atoms) and photon decoupling are related but not the same thing. With more precise definitions, they happen at slightly different times. Incidentially, the transition from radiation dominance to matter dominance also happens at about the same time.
Matter (electrons, protons, neutrons) is non-relativistic at decopuling and evolves as a thermal non-relativistic distribution with temperature

$$\frac{T}{T_{\mathrm{rec}}} = \frac{R_{\mathrm{rec}}^2}{R^2} \tag{7.162}$$

which corresponds to a temperature today of

$$T \approx 2 \cdot 10^{-3} K \tag{7.163}$$

But things are actually more complicated because once the temperature is low enough, and once there is now significant pressure from radiation, then matter forms gravitational bound states and a hierarchy of structures occurs (including ourselves). This is of course a non-equilibrium process. In contrast, the CMB is the most perfect thermal distribution ever observed. Since most of the entropy (and particle number) of the universe reside there (because photons are the only remaining relativistic particles), the description of the expanding universe as an adiabatic process is quite accurate. But certainly the CMB is not the most interesting or relevant part of the universe (also note that the energy is dominated by matter).

7.6.2 The neutrino background

Conservation of entropy implies for radiation:

$$S = \sum_i s_i V = \frac{2\pi^2}{45} g_{*,S} T^3 R^3 \sim g_{*,S} T^3 R^3 = \text{const.} \tag{7.164}$$

This means that the temperature goes like

$$T \sim g_{*,S}^{-\frac{1}{3}} R^{-1} \tag{7.165}$$

Note that this is different from the scaling of the distribution for a decoupled massless particle, $T \sim R^{-1}$. This difference matters when the number of effective massless degrees of freedom $g_{*,S}$ changes. This happens whenever the

temperature drops below the production scale of a massive particle at $m \approx T$. While the detailed behaviour at this scale might be complicated, we can compare the two asymtotic regimes. for $T \gg m$ the particle is ultrarelativistic and contributes to $g_{*,S}$. Once it is non-relativistic $T \ll m$ its distribution is exponentially suppressed and we remove it from $g_{*,S}$. Since entropy is conserved, this means that roughly all the entropy stored in the species which becomes non-relativistic is transferred to the remaining relativistic species. This slows down the reduction of the temperature.
A relavant example is the difference of the temperatures of the thermal photon and neutrino backgrounds. Neutrinos decouple at $T \approx 1$MeV, when electrons are still relativistc, while photons decouple at $T \approx 0.31$eV, where electrons are non-relativistic. The threshold for pair production at $T \approx 0.5$MeV

$$e + \overline{e} \leftrightarrow \gamma + \gamma \tag{7.166}$$

is between these two decoupling scales. Below the threshold electrons and positrons annihilate (up to the little matter surplus that we observe) and transfer their entropy to the photons.

1. $T > 1$MeV: Electrons, photons and neutrinos in equilibrium through electroweak interactions.

2. 1MeV $> T_1 >$ 0.5MeV. Electrons and photons in equilibrium through electromagnetic. Neutrinos decouple because weak interactions have become too rare, but distribution remains thermal, because neutrinos are effectively massless. Both temperatures scale like $T \sim S^{-1}$:

$$T_1 = T_{1,\gamma} = T_{1,\nu} \tag{7.167}$$

   Both systems have their own conserved entropies $S_{1,\gamma,e}$ = Entropy of photons and electrons, $S_{1,\nu}$ = entropy of neutrinos.

3. $T_2 <$ 0.5MeV. Electrons have annihilated (or become non-relativistic). Entropies are conserved:

$$S_{2,\gamma} = S_{1,\gamma,e} \ , \quad S_{2,\nu} = S_{1,\nu} \tag{7.168}$$

   The first equation implies

$$R_2^3 T_{2,\gamma}^3 g_{*,S,2} = R_1^3 T_1^3 g_{*,S,1} \tag{7.169}$$

   while the second gives

$$R_2^3 T_{2,\nu}^3 g_{*,S,1} = R_1^3 T_1^3 g_{*,S,1} \tag{7.170}$$

   For the neutrinos the distribution valid at neutrino decoupling continues to scale. They are not sensitive to the change in $g_{*,S}$. Therefore photons and neutrinos now have different temperatures $T_{2,\nu} = T_\nu$, $T_{2,\gamma} = T_\gamma$, related by

$$T_\gamma^3 g_{*,S,2} = T_\nu^3 g_{*,S,1}$$

$$\Rightarrow \quad \frac{T_\gamma}{T_\nu} = \left(\frac{g_{*,S,1}}{g_{*,S,2}}\right)^{\frac{1}{3}} \tag{7.171}$$

For 1MeV $> T_1 >$ 0.5MeV the relativistic particles in equlibrium are: electrons ($g = 2$), positrons ($g = 2$) and photons ($g = 2$):

$$g_{*,S,1} = 2 + \frac{7}{8} \cdot 2 \cdot 2 = \frac{11}{2} \tag{7.172}$$

If just photons remain relativistic:

$$g_{*,S,2} = 2 \tag{7.173}$$

so that

$$\frac{T_\gamma}{T_\nu} = \left(\frac{g_{*,S,1}}{g_{*,S,2}}\right)^{\frac{1}{3}} = \left(\frac{11}{4}\right)^{\frac{1}{3}} \tag{7.174}$$

and

$$T_\gamma \approx 2.73K \Rightarrow T_\nu \approx 1.96K \tag{7.175}$$

7.6.3 Particle densities

We can compare the particle, energy and entropy densities of various species. Since relativistic particles are present the entropy density of the universe is

$$s = \frac{2\pi^2}{25} g_{*,S} T^3 = \frac{2\pi^2}{25} g_{*,S} \left(\frac{T}{\hbar c}\right)^3 \tag{7.176}$$

If neutrinos are (sufficiently) massive, the only ultrarelativistic species are photons, so that $g_* = g_{*,S} = 2$. For three massless chiral neutrino families (as used in most available textbooks) this was instead:

$$g_* = 2\left(\frac{T_\gamma}{T_\gamma}\right)^4 + \frac{7}{8} \cdot 3 \cdot 2 \left(\frac{T_\nu}{T_\gamma}\right)^4 \approx 3.36 \tag{7.177}$$

$$g_{*,S} = 2\left(\frac{T_\gamma}{T_\gamma}\right)^3 + \frac{7}{8} \cdot 3 \cdot 2 \left(\frac{T_\nu}{T_\gamma}\right)^3 \approx 3.91 \tag{7.178}$$

The particle density of photons is

$$n_\gamma = 2 \cdot \frac{\zeta(3)}{\pi^2} T^3 \approx \frac{2.4}{\pi^2} T^3 \tag{7.179}$$

implying

$$\frac{s}{n_\gamma} \approx 1.8 g_{*,S} \tag{7.180}$$

The total number of baryons (non-relativistic matter) in the universe is assumed to be conserved and non-vanishing, so that baryon particle density n_B scales like R^{-3}. Then

$$\frac{n_B}{n_\gamma} = \frac{n_B}{s}\frac{s}{n_\gamma} =: B \cdot 1.8 g_{*,S} \tag{7.181}$$

The value of n_γ today is obtained by substituting the observed temperature $T_\gamma \approx 2.73K$. The density of baryons is estimated from the observation of luminous matter in the universe. This gives

$$\frac{n_B}{n_\gamma} = 10^{-8}\Omega h^2 \tag{7.182}$$

where Ω is the density function $\Omega = \frac{\rho}{\rho_c}$ $(\Omega \approx 1)$ and h is defined through

$$H(t_1) = 100hkm \cdot s^{-1}(Mpc)^{-1} \tag{7.183}$$

all at $t_0 =$ today. $h \approx 0.7$. Then

$$B \approx 1.4 \cdot 10^{-9}\Omega(t_0)h^2 \tag{7.184}$$

This illustrates that the particle density is dominated by photons (same with entropy). But due to the rest mass, the energy of the universe is dominated by matter (see below).
Some explicit values for illustration (values are taken from older text books and might not be completely accurate, but give correct orders of mangitude):
energy density of photons

$$\rho_\gamma = \frac{\pi^2}{30} g_*(t_1) T_{\gamma,1}^4 \approx 8 \cdot 10^{-34} g \cdot cm^{-3} \tag{7.185}$$

number density of photons

$$n_\gamma = 2\frac{\zeta(3)}{\pi^2} T_{\gamma,1}^3 = 422cm^{-3} \tag{7.186}$$

enropy density of photons

$$s_\gamma = \frac{2\pi^2}{45} g_{*,S}(t_1) T_{\gamma,1}^3 \approx 2970cm^{-3} \tag{7.187}$$

energy density for (chiral massless) neutrinos

$$\rho_\nu = 3 \cdot \frac{7}{8}\left(\frac{T_\nu}{T_\gamma}\right)^4 \rho_\gamma = \frac{21}{8}\left(\frac{4}{11}\right)^{\frac{4}{3}} \rho_\gamma \tag{7.188}$$

number density of neutrinos

$$n_\nu = \frac{9}{11} n_\gamma \approx 345cm^{-3} \tag{7.189}$$

entropy density for neutrinos

$$s_\nu = \frac{9}{11} s_\gamma \tag{7.190}$$

7.6.4 Transition from radiation to matter dominance

Let t_E be the time where the energy densities of radiation and matter become equal.

$$\rho_R(t_E) = \rho_M(t_E) \tag{7.191}$$

The radiation (photon) density can be related to the one observed today:

$$\rho_R(t_E) = a_0 T_E^4 = a_0 \frac{R_0^4}{R_E^4} T_0^4 \tag{7.192}$$

where $t_0 =$ today, and we know from the CMB that $T_0 \approx 2.75K$. Thus

$$\rho_{\gamma,0} = \frac{\pi^2}{30} g_* T_0^4 \approx 8 \cdot 10^{-34} g\ cm^{-3} \tag{7.193}$$

The matter density evolves like

$$\rho_M(t_E) = \rho_M(t_0) \frac{R_0^3}{R_E^3} \tag{7.194}$$

The value today can be estimated from observing luminous matter:

$$\rho_{M,0} = 1.8 \cdot 10^{-29} \Omega_0 h^2 g\ cm^{-3} \tag{7.195}$$

where Ω_0 is the (full) density function for today and h the empirical factor in the Hubble constant H_0. Plugging in, we have

$$\frac{R_1}{R_E} = \frac{\rho_{M,0}}{a_0 T_0^4} \approx \frac{1.8 \cdot 10^{-29} \Omega_0 h^2 g\ cm^{-3}}{8 \cdot 10^{-34} g\ cm^{-3}} \approx 2.25 \cdot 10^4 \Omega_0 h^2 = 1 + z_E \tag{7.196}$$

In the last step we related the expansion rate to the redshift. Temperature and time:

$$T_E = T(1 + z_E) = 5.5 \Omega_0 h^2 eV \tag{7.197}$$

$$t_E \simeq \frac{2}{3} H_0^{-1} \Omega_0^{-\frac{1}{2}} (1 + z_E)^{-\frac{3}{2}} \approx 1.4 \cdot 10^3 (\Omega_0 h^2)^{-2} y \tag{7.198}$$

If one evaluates t_E exactly (instead of using the asymptotic form for matter dominated time evolution), then

$$t_E = 0.39 H_0^{-1} \Omega_0^{-1/2} (1 + z_E)^{-3/2}\ y \tag{7.199}$$

7.6.5 Recombination

'Recombination' refers to the formation of atoms in the early universe. It is a misnomer, because this is presumably the first time where atoms formed. At recombination time ($\simeq 10^5 y$)the only (relavant) stable particles are protons, neutrons, electrons and photons. Free neutrons have a life time of a few minutes, but neutrons can survive as constituents of nuclei. According to the theory of primordial nucleosynthesis, the only relevant nuecleus at recombination time are ^4He, which makes up 10 percent of the baryonic particle density (the rest is protons). We will ignore the helium for simplicity. and we ignore Helium for simplicity. Then the relevant process is

$$p + e \leftrightarrow H + \gamma \tag{7.200}$$

Neutrality of matter and baryon number conservation implies

$$n_p = n_e \ , \quad n_B = n_p + n_H \tag{7.201}$$

The relevant energy scale is the ionization energy of hydrogen,

$$E_{ion} = m_e + m_p - m_H \approx 13.6eV \tag{7.202}$$

All matter particles are non-relativistic at the relavant temperature:

$$m_i \gg T \tag{7.203}$$

The number density distributions are non-relativistic

$$n_i = g_i \left(\frac{m_i T}{2\pi}\right)^{\frac{3}{2}} \exp\left(\frac{\Phi_i - m_i}{T}\right) \tag{7.204}$$

In chemical equilibrium (and using that photons are relativistic) we have

$$\Phi_p + \Phi_e = \Phi_H \tag{7.205}$$

Using these equations

$$\begin{aligned}
n_H &= g_H \left(\frac{m_H T}{2\pi}\right)^{\frac{3}{2}} \exp\left(\phi_H - m_H\right) \\
&= g_H \left(\frac{m_H T}{2\pi}\right)^{\frac{3}{2}} \exp\left(\phi_e + \phi_p - m_H\right) \\
n_{e,p} &= g_{e,p} \left(\frac{m_{e,p} T}{2\pi}\right)^{\frac{3}{2}} \exp\left(\phi_{e,p} - m_{e,p}\right) \\
\Rightarrow \ \exp\left(\tfrac{\phi_{e,p}}{T}\right) &= \frac{n_{e,p}}{g_{e,p}} \left(\frac{2\pi}{m_{e,p} T}\right)^{\frac{3}{2}} \exp\left(\frac{m_{e,p}}{T}\right) \\
\Rightarrow n_H &= \frac{g_H}{g_e g_p} \left(\frac{2\pi m_H}{m_e m_p T}\right)^{\frac{3}{2}} \exp\left(\frac{E_{\rm ion}}{T}\right) \\
&\approx \frac{g_H}{g_e g_p} n_e n_p \left(\frac{2\pi}{m_e T}\right)^{\frac{3}{2}} \exp\left(\frac{E_{ion}}{T}\right)
\end{aligned} \tag{7.206}$$

(In the factor before the exponent we used $m_p \approx m_H$.)

The fractional ionization rate

$$X_e = \frac{n_p}{n_B} = \frac{n_p}{\eta n_\gamma} \tag{7.207}$$

where $\eta := \frac{n_B}{n_\gamma}$. Compute

$$\frac{1 - X_e}{X_e^2} = \frac{n_H \eta n_\gamma}{n_p^2} = \eta \frac{4\sqrt{2}\zeta(3)}{\sqrt{\pi}} \left(\frac{T}{m_e}\right)^{\frac{3}{2}} \exp\left(\frac{E_{ion}}{T}\right) \tag{7.208}$$

(We used $g_H = 4$, $g_e = g_p = 2$.)
This is the so-called Saha equation which determines the equilibrium ionization fraction.
Use values

$$\eta = 2.68 \cdot 10^{-8} \Omega_B h^2 \tag{7.209}$$

(where Ω_B is the baryonic matter density in critical units, while Ω_0 is the total matter density in critical units)

$$T = (1 + z)2.73K \tag{7.210}$$

With $0.1 < \Omega_B h^2 < 1$ and defining recombination by $X_e = 0.1$ recombination occurs at

$$1200 < 1 + z < 1400 \tag{7.211}$$

Taking $1 + z = 1300$,

$$T_{rec} = T_0(1 + z_{rec}) = 3375K = 0.308eV \neq 13.6eV = E_{ion} \tag{7.212}$$

The recombination temperature is one order of magnitude smaller than the ionization temperature, due to the large entropy factor. At recombination time we already have matter dominance. Time

$$t_{rec} = \frac{2}{3} H_0^{-1} \Omega_0^{-\frac{1}{2}} (1 + z_{rec})^{-\frac{3}{2}} \approx 4.39 \cdot 10^{12} (\Omega_0 h^2)^{-\frac{1}{2}} s \tag{7.213}$$

A more elaborate analysis shows that the use of the equilibrium ionization is justified for $1 + z > 1100$. (For smaller redshift there is a residual ionization. This is a non-equilibrium process.)
Using that $n_p = n_e$, equilibrium ionization provides us with the density of free electrons, which (potentially) keep matter and radiation in equilibrium.

$$n_e = X_e n_B = X_e \eta n_\gamma = X_e(\Omega_B h^2)(1 + z)^3 1.13 \cdot 10^{-5} \mathrm{cm}^{-3} \tag{7.214}$$

Equilibrium is maintained until the interaction rate Γ becomes smaller than the expansion rate H. The threshold of decoupling is

$$\Gamma \simeq H \tag{7.215}$$

where

$$\Gamma = n_e \sigma_T \tag{7.216}$$

The relevant cross section for the relevant Thompson scattering is

$$\sigma_T = 6.65 \cdot 10^{-25}\mathrm{cm}^2 \tag{7.217}$$

The redshift at decoupling z_{dec} depends on Ω_0 (through H) and Ω_B (through n_e). The range is

$$1100 \leq 1 + z_{dec} \leq 1200 \tag{7.218}$$

For $z_{dec} = 1100$ the decoupling temperature is

$$T_{dec} = T_0(1 + z_{dec}) = 3030K = 0.26eV \tag{7.219}$$

and the decoupling time is

$$t_{dec} = \frac{2}{3}H_0^{-1}\Omega_0^{-1/2}(1 + z_{dec})^{-3/2} = 5.64 \cdot 10^{12}(\Omega_0 h^2)^{-1/2}s \tag{7.220}$$

(using the world age for matter dominance).
Reminder: We have not adusted the times and redshifts for decoupling, recombination, etc. to the most recent data. Results are roughly accurate, but with recent data precision has grown. We hope to improve on this in a later version, but also encourage the reader to investigate the necessary adjustments herself/himself.

Chapter 8

Inflation

Appendix A

Some formulae

A.1 Units and constants

Boltzmann constant

$$k_B = 1.38 \cdot 10^{-23} \frac{J}{K} = 8.62 \cdot 10^{-5} \frac{eV}{K} \tag{A.1}$$

In units where $k_B = 1$:

$$1K = 8.62 \cdot 10^{-5} eV \ , \quad 1eV = 1.16 \cdot 10^{4} K \tag{A.2}$$

A.2 Cosmological formulae

Temperature–redshift relation

$$1 + z = \frac{T(t)}{T_0} = \frac{T}{2.73K} \tag{A.3}$$

Temperature–age relation(??)

$$t \approx 2 \left(\frac{10^{10}}{T[K]} \right) s \tag{A.4}$$

(does not seem to be very accurate).

A.3 Some integrals and sums

All sums and integrals can be found in [24].

A.3.1 Sums related to the Riemann ζ-function

For $s \in \mathbb{C}$ with $\mathrm{Re}\, s > 0$, the Riemann ζ-function is defined by the sum

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \tag{A.5}$$

By analytic continuation, one obtains a holomorphic function on $\mathbb{C} - \{1\}$ with a simple pole at $s = 1$.
Some special values:

$$\zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \tag{A.6}$$

$$\zeta(4) = \sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{60} \tag{A.7}$$

The sums can be evaluated by contour integration techniques, [23] appendix A 11.
At $s = 3, 5, 7, \ldots$, the value $\zeta(s)$ is tranzcendental. Approximate values can be computed numerically:

$$\zeta(3) = 1.20206\ldots \tag{A.8}$$

Some related alternating sums can be computed by contour integration techniques

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2} = \frac{\pi^2}{12} = \frac{1}{2}\zeta(2) \tag{A.9}$$

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^4} = \frac{7\pi^4}{720} = \frac{7}{8}\zeta(4) \tag{A.10}$$

Alternatively these sums can be related directly to $\zeta(2k)$, $k = 1, 2, \ldots$, by a trick explained below. The same trick can be used to relate other alternating sums to transcendental values of ζ:

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^3} == \frac{3}{4}\zeta(3) \tag{A.11}$$

A.3.2 Integrals related to the Γ-function

$$\int_0^{\infty} dx\, x^m e^{-x} = \Gamma(m+1) = m! \tag{A.12}$$

for $m = 0, 1, 2, \ldots$.
This is a special case of the more general integral

$$\int_0^{\infty} dx\, x^m e^{-\alpha x} = \frac{\Gamma(m+1)}{\alpha^{m+1}} \tag{A.13}$$

for real m, α with $\alpha > 0$ and $m > -1$.

A.3.3 Integrals related to the Gaussian integral

$$\int_0^\infty dx\, x^m e^{-\alpha x^2} = \frac{\Gamma(\frac{m+1}{2})}{2\alpha^{\frac{n+1}{2}}} \tag{A.14}$$

for real m, α with $\alpha > 0$ and $m > -1$. For even m these can be obtained from the Gaussian integral

$$\int_0^\infty dx\, e^{-\alpha x^2} = \frac{\sqrt{\pi}}{2\alpha} \tag{A.15}$$

by differentiation with respect to α:

$$\int_0^\infty dx\, x^2 e^{-\alpha x^2} = -\frac{d}{d\alpha}\frac{\sqrt{\pi}}{2\alpha} = \frac{\sqrt{\pi}}{4\alpha^3} \tag{A.16}$$

A.3.4 Integrals of the form $\int_0^\infty dx \frac{x^m}{e^x - \varepsilon}$

Thermodynamic densities involve integrals of the form

$$\int_0^\infty dx \frac{x^m}{e^x - \varepsilon} \tag{A.17}$$

for $\varepsilon = +1$ (Bose-Einstein statistics) and $\varepsilon = -1$ (Fermi-Dirac statistics).

Consider first $\varepsilon = +1$. Expand the integrand:

$$\int_0^\infty dx \frac{x^m}{e^x - 1} = \sum_{n=1}^{\infty} \int_0^\infty dx x^m e^{-nx} \tag{A.18}$$

$$= \sum_{n=1}^{\infty} \frac{1}{n^{m+1}} \int_0^\infty dy y^m e^{-y} \tag{A.19}$$

$$= \zeta(m+1)\Gamma(m+1) \tag{A.20}$$

For $\varepsilon = -1$, one can split the integral into two integrals of the previous type

$$\int_0^\infty dx \frac{x^m}{e^x + 1} = \int_0^\infty dx \frac{x^m}{e^x - 1} - 2\int_0^\infty dx \frac{x^m}{e^{2x} - 1} \tag{A.21}$$

$$= \left(1 - \tfrac{2}{2^{m+1}}\right) \int_0^\infty dx \frac{x^m}{e^x - 1} \tag{A.22}$$

$$= \tfrac{2^m - 1}{2^m} \zeta(m+1)\Gamma(m+1) \tag{A.23}$$

Since

$$\int_0^\infty dx \frac{x^m}{e^x + 1} = \sum_{n=1}^{\infty} \int_0^\infty dx x^m (-1)^{n+1} e^{-nx} \tag{A.24}$$

$$= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^{m+1}} \int_0^\infty dy y^m e^{-y} \tag{A.25}$$

$$= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^{m+1}} \Gamma(m+1) \tag{A.26}$$

we obtain a relation between infinite sums

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^s} = \frac{2^s - 1}{2^s} \sum_{n=1}^{\infty} \frac{1}{n^s} \tag{A.27}$$

By analytic conitinuation this becomes a relation between the Riemann ζ function $\zeta =: \zeta_1$ and a modified ζ-function ζ_{-1}, where

$$\zeta_\varepsilon(s) = \sum_{n=1}^{\infty} \frac{\varepsilon^{n+1}}{n^s} \tag{A.28}$$

for $\varepsilon = \pm 1$.
Then we can write our thermodynamic integral in compact form

$$\int_0^\infty dx \frac{x^m}{e^x - \varepsilon} = \zeta_\varepsilon(m+1)\Gamma(m+1) \tag{A.29}$$

Bibliography

[1] L.D. Landau and E.M. Lifshitz, Vol 2

[2] L.D. Landau and E.M. Lifshitz, Vol 6 5.1.7

[3] J.A. Peacock, Cosmological Physics, CUP. 6, 2.3, 5.1.7, 7.5.2

[4] R. d'Inverno, Einführung in die Relativitätstheorie, VCH. 5.1.7

[5] S.W. Hawking and G.F.R. Ellies, The large scale structure of space-time, CUP 5.1.7, 5.1.7

[6] C.W. Misner, K.S. Thorne and J.A. Wheeler, Gravitation, Freeman. 5.1.3, 5.1.7

[7] T.M. Davis and C.H. Lineweaver, Expanding Confusion: common misconceptions of cosmological horizons and the superluminal expansion of the universe, astro-ph/0310808.

[8] M. Visser and C. Barcelo, Energy conditions and their cosmological implications, gr-qc/0001099

[9] J.P. Ostriker and T. Souradeep, The current status of observational cosmology, astro-ph/0409131. 2.2

[10] R.U. Sexl and H.K. Urbantke, Relativity, Groups, Particles. Springer. 2.2, 4.9.3, 6.3

[11] R.U. Sexl and H.K. Urbantke, Gravitation und Kosmologie, BI Wissenschaftsverlag. 6.3

[12] A.R. Liddle and D.H. Lyth, Cosmological Inflation and Large-Scale structure, CUP. 6

[13] H. Goenner, Einführung in die Kosmologie, Spektrum Verlag. 2, 3, 2.3, 2.4.1, 2.4.2, 1, 5.2.7, 5.2.8, 7.3

[14] A.G. Riess et al, The Farthest Known Supernova: support for an accelerating universe and a glimpse of the epoch of deceleration, astro-ph/0104455. 5

[15] E.K. Akhmedov, Neutrino physics, hep-ph/0001264. 1

[16] K. Grotz und H.V. Klapdor, Die schwache Wechselwirkung in Kern-, Teilchen- und Astrophysik, Teubner. 1

[17] H. Stephani, Titel, VEB Deutscher Verlag der Wissenschaften, 1977. 4.9.3, 4.9.3, 4.9.3

[18] General Relativity, Lecture Notes, Univ. Hamburg, SS 1982. DESY T-82-03. 4.9.3, 4.9.3

[19] H.H. Voigt, Abriß der Astronomie, BI Wissenschaftsverlag, 1980. 2.3

[20] Lexikon der Astronomie, Spectrum Akademischer Verlag, 1999. 2.3

[21] V.P. Frolov and I.D. Novikov, Black Hole Physics, Kluwer Academic Publishers, 1998. 2.3

[22] C. L. Bennett, First year Wilkinson Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results, astro-ph/0302207. 2.6

[23] F. Reif, Statistische Physik und Theorie der Wärme, de Gruyter, 1987. A.3.1

[24] E. Zeidler (Hrg.), Teubner-Taschenbuch der Mathematik, Teubner, 1996. 5.2.8, A.3

[25] E.W. Kolb and M. S. Turner, the Early Universe, Addison Wesley, 1990. 5.2.8, 5.2.8, 5.2.8, 5.2.8